USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No.

A Novel Decision Support Tool to Develop Link Driving Schedules for Moves

By

Satish Ukkusuri
Professor of Civil Engineering
Purdue University
sukkusur@purdue.edu

DISCLAIMER

Funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT12-G-UTC05 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program.

A Novel Decision Support Tool to Develop Link Driving Schedule for MOVES

Introduction

A system or user level strategy that aims to reduce emissions from transportation networks requires a rigorous assessment of emissions inventory for the system to justify its effectiveness. It is important to estimate the total emissions for a transportation network before and after the implementation of a particular policy. For instance, a traffic signal control scheme that is optimized for environmental goal is expected to cause less emissions compared with the scheme without environmental goal. This research proposes a novel technique to find the representative vehicle trajectories and the corresponding LDS for links on transportation networks. The technique uses the dynamic time warping distance as the similarity measure in clustering which is more appropriate for curve alignment compared with Euclidean distances and its variants which is more common in the literature.

Findings

- The HC-DTW technique provides higher accuracy compared with average speed technique as we see from our seven test cases.
- The error percentage for PM_{10} is found to very high in most cases. The results for $PM_{2.5}$ have similar trends and are not reported here for the sake of brevity.
- The number of links in a cluster affects the accuracy of the estimation. As the number of vehicle increases in a cluster, the degree of similarity (closeness) decreases. Although the average similarity remains same, some details are lost. Further, estimation with a very high volume traffic adds the accumulated error. This implies that the when the number of trajectories are reasonably high the analyst needs to carefully decide the number of clusters to include as input to MOVES.
- Except PM_{10}, the error percentage ranges from 1% to 10% for all other pollutants. Case 2 and case 6 exhibit higher errors for NO_x.

Recommendations

- Machine learning algorithms are effective techniques to estimate emissions at the city wide level
- The developed tool is an effective add-on to MOVES to estimate emissions effectively
- HC-DTW is most effective when the variation in vehicular activities on a link is high which is typically found during peak hour conditions and in urban areas (section 5.1.1). If the congestion level is low, and the vehicular activities are similar, average speed technique provides estimation with reasonable accuracy.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1. BACKGROUND

A system or user level strategy that aims to reduce emissions from transportation networks requires a rigorous assessment of emissions inventory for the system to justify its effectiveness. It is important to estimate the total emissions for a transportation network before and after the implementation of a particular policy. For instance, a traffic signal control scheme that is optimized for environmental goal is expected to cause less emissions compared with the scheme without environmental goal. To assess the benefits of the new signal control scheme one needs to compare the emissions level with the base case. Further, air quality assessment is a requirement for the local agencies for conformity and funding decisions. The Clean Air Act section 176(c) requires transportation conformity to ensure that federally supported highway and transit project activities support and move towards the state air quality implementation plan (SIP). Federal Highway Administration (FHWA) requires transportation conformity studies to ensure that federal funded and approved projects are consistent with air quality goals set by National Ambient Air Quality Standards (NAAQS).

Integrated traffic-emissions simulator framework is gaining much attention recently as a primary tool to assess the emissions reduction policies. The U.S. Environmental Protection Agency (EPA) has regulated to use MOVES to develop state implementation plans (SIP) and regional or project-level transportation conformity analyses. MOVES is a modal-based emissions estimator that accounts for vehicle operating modes defined by factors like speed, acceleration, road grade, curvature, and so on. MOVES has the ability to include alternative types of fuel and different type of vehicles as well. Analyses at different scales including regional, state, and project level (e.g., small road network at county level) can be done with MOVES.

1.1 Emissions estimation using EPA-MOVES

Integration of MOVES with traffic simulator can be outlined as an input-output process. The second-by-second vehicular activities from traffic simulation serve as input for MOVES and the emissions inventory for a transportation network can be estimated (Mahut and Florian, 2010; Lin et al., 2011; Xie et al., 2012; Hao et al., 2012).The input from traffic simulators can be any of the following formats: (a) average speeds for the links in the network, (b) Link Driving Schedule (LDS) for each link of the network.

LDS is time dependent speed profile of a link (generally done for a representative vehicle or by means of sampling), and (c) Operating mode distribution of vehicles on the link. While average speed is commonly used in practice, operating mode distribution and LDS can take the advantages regarding vehicular activity data and dynamic capability of MOVES to report time dependent emissions (EPA, 2012a).

1.2 Challenges wih urban signalized networks

Input data for MOVES for the analysis of signalized intersections require more details and careful selection of vehicular activities. EPA suggests to use a series of LDS to represent cruise, acceleration, idle, and deceleration behavior in a congested intersection. Also notions of approach and departure link should be incorporated. Two challenges with this link segmentation technique are: (a) the overlapping of cruise, idle, acceleration, deceleration, and idle zone (e.g., the vehicular activity 100 ft. from the stop line at red and green phases), (b) Finding a single representative LDS that features all the vehicles within that segment for the analysis period (e.g., one may need to pick one LDS from 400 vehicles for a single hour analysis). Op-mode distribution requires computing fractions of the eet travel times spent in each mode of operations. The computation is based on ratios of mode travel times to total travel times and significant post processing is required (EPA, 2012a).

An alternative to the link segmentation technique is to input the second-by-second trajectory of the vehicles. In an integrated traffic-emissions framework LDS is commonly used as input for MOVES because trajectories are relatively easy to obtain and can be converted into LDS for links. Moreover, it allows to overcome the first limitation of link segmentation technique. We do not have to separately model idling, cruise, acceleration, and deceleration because vehicular trajectories accommodates all possible activities. However,the challenge is to find the representative trajectory for a link that may be traversed by thousands of vehicles during the analysis period. Sampling techniques with carefully designed selection criteria can be used to reduce the size of the problem. However the computation becomes significantly expensive as the size of the problem grows. A sample of 5000 vehicles for 150,000 original data points can take multiple hours (if not days). This makes the trajectory approach inefficient in terms of computation and obviously accuracy is compromised as the sample size is reduced.

1.3 Research goals

This research proposes a novel technique to find the representative vehicle trajectories and the corresponding LDS for links on transportation networks. The technique uses the dynamic time warping distance as the similarity measure in clustering which is more appropriate for curve alignment compared with Euclidean distances and its variants which is more common in the literature. A recent approach by Chamberlin et al. (2011) applies K-means clustering with Euclidean distances. Two major limitations exist with this study. First, the Euclidean distance is not an appropriate similarity measure for time series clustering. Euclidean distance measures cannot account for this shift in time phase. Euclidean distance fails to provide a correct measure of similarity between two sequences because of its high sensitivity to changes in the time axis (Chu et al., 2002; Keogh and Pazzani, 1999). Figure 1(a) illustrates the benefits of using DTW measure for time-series alignment. As described in Keogh and Ratanamahatana (Keogh and Ratanamahatana, 2005), the two curves have almost similar shape except they are not aligned in the horizontal axis. Euclidean distance only measures the difference between the same two points in the horizontal axis and obviously the dissimilarity will be higher. On the other hand, DTW measures are non-linear and captures the shift in horizontal axis. Second, K-means algorithm requires the number of clusters to be predefined.

Our proposed technique overcomes both limitations. First, we use a dynamic time warping (DTW) measure instead of Euclidean distance. The advantage DTW offers is that it can align two times series having similar shape but not aligned in time. DTW allows elastic shifting of the time axis to identify similar shapes with different phases. Figure 1(b) illustrates the DTW similarity for two trajectories with different lengths. The trajectories are overall similar but shifted in time.

We applied agglomerative hierarchical clustering that does not require the number of clusters as input. The number of clusters is determined using inconsistency coefficient measure and also it is possible to assign maximum number of clusters. Finally, our proposed technique reduces the computation requirement significantly with negligible compromise in the accuracy of estimation.

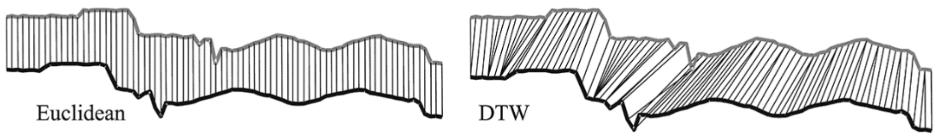
The goals of this research are as follows:

(a) To develop a similarity based clustering technique that finds the representative trajectories accounting for heterogeneous vehicular activities on the link,

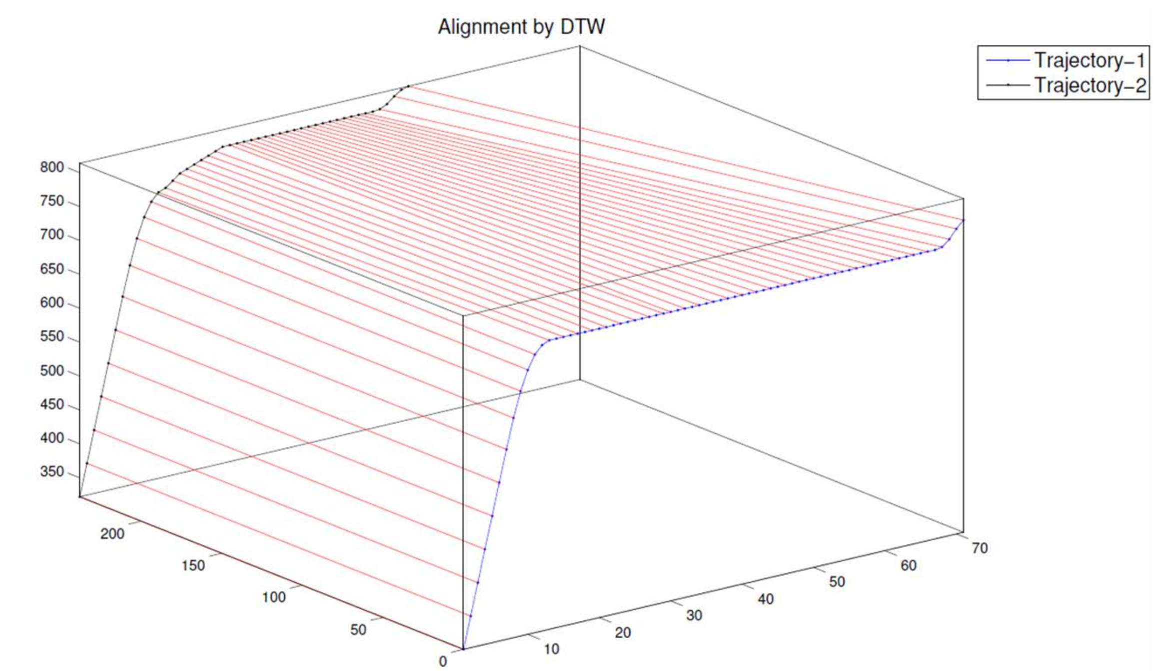
(b) To find optimal sets of Link Driving schedule (LDS) using representative trajectories on a link that can be input for MOVES to estimate emissions,

(c) To demonstrate the applicability of the technique in terms of accuracy and computational efficiency for the analyses of network with signalized intersections.

The rest of the paper is organized as follows: chapter 2 provides an overview of MOVES2010, chapter 3 describes the integration framework, chapter 4 dynamic time warping based clustering technique in details, chapter 5 demonstrates the applicability of our proposed technique, and finally we discuss the limitations and future research direction in chapter 6.



(a) Comparing Euclidean distance vs. dynamic time warping measure(Keogh and Ratanamahatana, 2005)



(b) Alignment by DTW

Figure 1. Advantages of using Dynamic time warping measure

CHAPTER 2. EPA REGULATED ESTIMATION TOOL: MOVES2010

MOVES (MOtor Vehicle Emissions Simulator) developed by the U.S. EPA was first released in December 2009. In March 2010, MOVES was officially approved for use in state implementation plans (SIP) and transportation conformity analysis EPA (2009) outside California. We have used MOVES2010b in this research to estimate emissions. We will use the term "MOVES" throughout the manuscript that refers to MOVES2010b version. MOVES can be used both as an inventory model providing estimation of total emissions for a network or region or even state) and as an emissions rate model. MOVES is one of the micro-scale vehicle emissions simulator that uses instantaneous (e.g., second-by-second) operations of individual vehicles on the road (Liu and Barth, 2012; EPA, 2012a). Among the few microlevel emission models, International Vehicle Emission (IVE) model (Lents and Davis, 2004) and Comprehensive Modal Emissions Model (CMEM) (Barth et al., 2000; Barth et al., 2004) are well known for estimating vehicular emission. MOVES is one of the most sophisticated emissions models that can be applied at di_erent modeling scales: from the micro-scale (project level, e.g., parking lot) to the macro-scale, where national-scale inventories are being generated for precursor, criteria, and greenhouse pollutants from on-road mobile sources (Liu and Barth, 2012).

2.1 Estimation Method in MOVES

The emission estimation in MOVES is primarily based on power demand of the vehicles. Two categorical bins are de_ned in the MOVES database: source bin and operating mode (op-mode) bin. The source bin is defined based on Vehicle characteristics (type of fuel, engine type, make and year, loaded weight, and engine size). The Operating mode bins are defined based on the second-by-second vehicle characteristics (idling, accelerating, cruising, decelerating, and so on). Vehicle Specific Power (VSP) is used as the measure of the power demand Jimenez-Palacios (1999) placed on a vehicle under various driving modes. VSP is defined as the power demand on the engine per unit of vehicle mass to surmount the inertial acceleration (power demand), rolling resistance, road grade, and aerodynamic drag (Jimenez-Palacios, 1999; Frey et al., 2010). Studies (Bapat and Gao, 2010; Huai et al., 2005; Liu and Barth, 2012) show direct correlation between VSP and vehicular emission. IVE Barth et al. (2004), and CMEM Barth et al. (2004) also use VSP based approach (Lents and Davis, 2004) to estimate emission from road networks.

VSP is differentiated based on driving cycles and vehicle characteristics. Other than the meteorological factors, vehicle characteristics highly impact the emission estimation. Also, the aerodynamic drag varies by vehicle size, type, and loads. For instance, the aerodynamic drag for a compact car and for a full size car will be quite different (Frey et al., 2010). Detail description on the emission estimation method can be found in the MOVES user guide (EPA, 2012a) and EPA guidelines for GHG and energy estimation from road networks (EPA, 2012b).

2.2 Datat input for MOVES

MOVES requires several traffic network related attributes for emissions estimation. These include

- Type of links (freeway, arterial, parking lot, truck terminal, and so on) based on the type of access (restricted vs. unrestricted)
- Length of link, grade (slope), and average speed of vehicles on the link for the analysis period.
- Special treatment is applied for the analysis of signalized intersections. Generally, the link is divided into several segments to account for the stopping of the vehicles at red.

Currently, MOVES has five categories of road types. The default database contains: a) Rural roads with restricted and unrestricted access, b) Urban roads with restricted and unrestricted access, c) Off-network (this is primarily for the extended idle process in parking lots or truck terminals). Note that, the selected road type may or may not exist in the geographical bounds of selected county and MOVES only computes results for the existing road types for that county.

The time span including the year, month, day (weekday or weekend), and starting-ending hours needs to be defined precisely. The finest resolution is one hour in MOVES. For project level analysis, any county in the U.S. can be chosen for analysis. Further, one can customize the county level parameters using the option custom domain. The input data include temperature, wind pressure and so on (see MOVES user guide (EPA, 2012a))

CHAPTER 3. INTEGRATION WITH TRAFFIC SIMULATION

The recent advances in the emission modeling allow to compute emissions accounting for second-by-second operation characteristics of a vehicle on a road segment (EPA, 2012a; Lin et al., 2011).The instantaneous emission models are useful for temporal and spatial analysis of different policies and emission reduction strategies at fine-grained levels. These models overcome the under-estimation problem associated with the air quality models that assume emissions are evenly distributed along a road section (Boulter and McCrae, 2007). Now, micro-scale models like MOVES require detailed and precise information regarding vehicle operation and location. Without a detail representation of vehicular activities the expected accuracy level will be low. On the other hand, collecting precise data on instantaneous vehicle operations for even a small network is tedious, expensive, and time consuming. An obvious solution to this problem is the use of micro level traffic simulation model that can effectively produce all the required inputs with desired details for the instantaneous emissions model. Figure 2 shows the general framework to integrate EPA-MOVES and a traffic simulator.

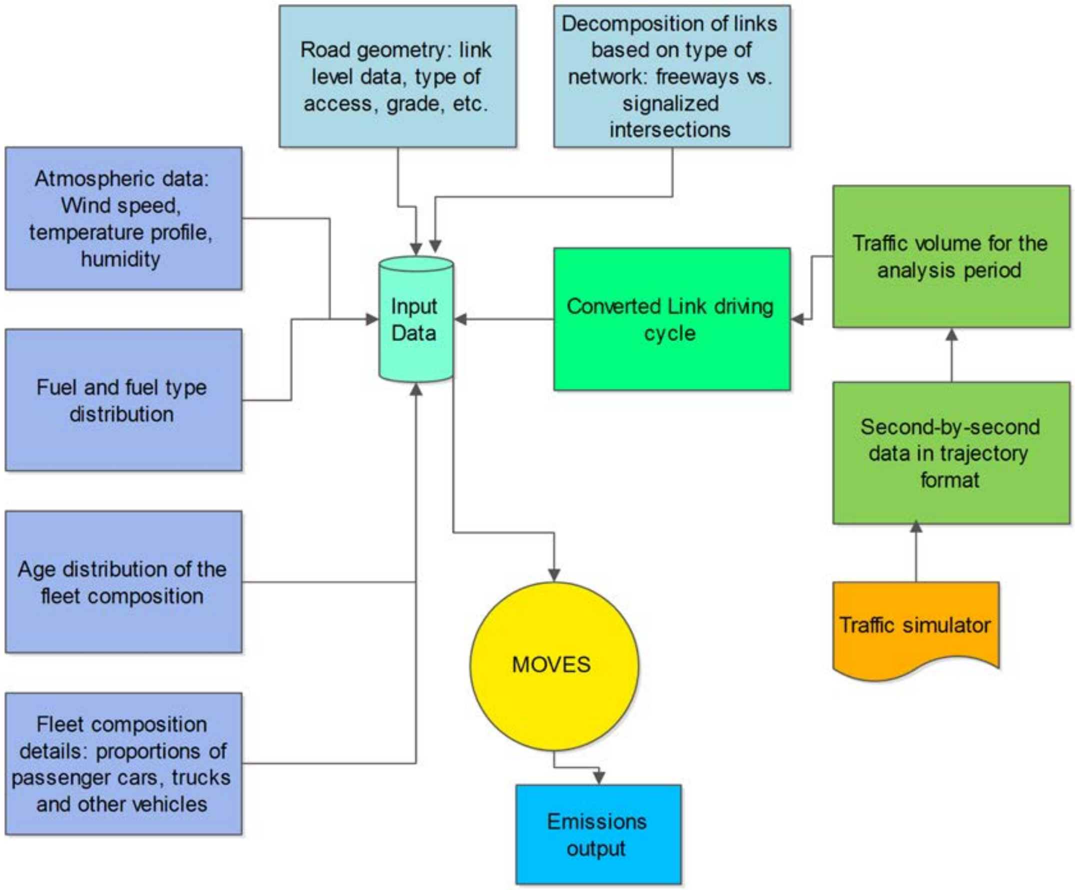


Figure 2. General framework of integrated traffic simulation and EPA-MOVES

CHAPTER 4. METHODOLOGY TO DEVELOP LINK DRIVING SCHEDULES

This chapter provides the details on developing link driving schedules for EPA-MOVES using the trajectories produced by a traffic simulator. The proposed methodology applies hierarchical clustering technique with dynamic time warping (DTW) (dis)similarity measures. Details on the methodology can be found in (Everitt et al., 2001). Figure 3 shows the steps of the technique proposed in this research. Hereafter we refer to our clustering technique as hierarchical clustering with DTW or HC-DTW technique. The inputs are the vehicular trajectories of all links in the network. The DTW algorithm provides the shortest DTW path distance for any two trajectories. For a link we compute the DTW distances for all trajectories to create a cost matrix. Each element in the matrix is the DTW measure with all other trajectories for the link. This matrix acts as the distance matrix for the hierarchical clustering. Note that we do not use k-means method because the objective is curve alignment based on similarity measures. The k-means algorithms are more suitable for density based clustering where the focus is more on the location.

Unlike k-means algorithm, hierarchical clustering does not require the number of clusters as an input. One can follow either agglomerative or divisive method to obtain hierarchical clusters. Both approaches provides the clusters and can be visualized using a dendrogram. This study follows agglomerative method. The clustering tree starts with all elements condensed in a single cluster on the top and each element a separate cluster at the bottom. One particular challenge is to find the optimal number of clusters. In other words, where should we cut the clustering tree? This research follows two approaches to cut the clustering tree: (a) inconsistency coefficients, (b) maximum number of clusters allowed. After obtaining the clusters, a prototype from each cluster is picked. The prototype trajectory is the member of the cluster that minimizes the distance with all other members within that cluster. The prototype trajectory is then converted into LDS by computing the instantaneous speed. The next subsections provide details on the DTW and clustering technique.

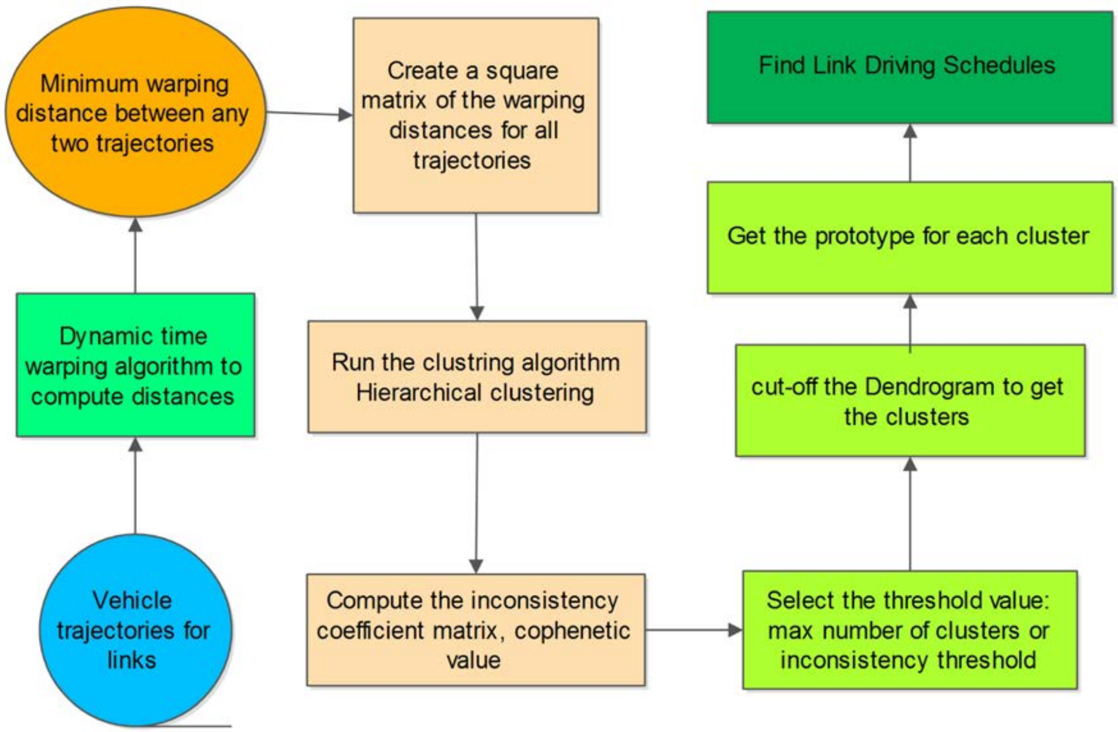


Figure 3. Overview of the Link Driving Schedule (LDS) finding methodology

4.1 Computing DTW measures

Dynamic time warping (DTW) (Sakoe and Chiba, 1978; Berndt and Clifford, 1994) is a similarity measure that can be applied for unsupervised clustering of time series. DTW finds an optimal alignment between any two time dependent sequences. The optimal alignment (also known as the minimum distance warp path) is determined by assigning successive values of one sequence to a single value of the other. This enables DTW to find alignment even when the time series are of different lengths.

Consider two sequences *u* and *v* with lengths *m* and *n* respectively. The seqeunces are: $u:\{u_1,u_2,\cdots,u_i,\cdots,u_m\}$ and $v:\{v_1,v_2,\cdots,v_j,\cdots,v_n\}$. A $n \times m$ grid can be constructed where each element in the grid represents an alignment between any two objects from *u* and *v*. A warping path W^p is a sequence of elements from this grid ($W^p = w_1, w_2, \cdots, w_k, \cdots w_K$). Each element w_k is defined with element index from *u* and *v*. The sequence *W* maps the elements of the *u* and *v*. Classical DTW path has several constraints:

- Boundary conditions: The warping path starts at first point of both sequences and ends in last point of both sequences. Mathematically, $w_1 = (1,1), w_K = (n,m)$
- Continuity: The warping path is restricted to the adjacent cells in all directions. Consider two sequential elements $w_k = (u,v)$ and $w_{k+1} = (u',v')$ in *W*. Now the conditions $u' - u \leq 1, v' - v \leq 1$ must hold.
- Monotonicity: The points in *W* are monotonically spaced in time.

Exponentially high number of warping paths are possible that satisfy the above conditions. DTW algorithm finds the path with minimal cost (distance).

$$DTW(u,v) = \min(1/K)\sqrt{\sum_{k=1}^{K} w_k} \tag{1}$$

Denote $\gamma(i,j)$ as a cumulative distance defined as:

$$\gamma(i,j) = d(u_i,u_j) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\} \tag{2}$$

Dynamic programming can be used to find this recurrence equation. Details on the algorithm can be found in Kruskall and Liberman (1983).

Figure 4 shows an example of computing DTW minimum path for two vehicular trajectories of different lengths (45 seconds vs. 70 seconds). We have used a publicly available machine learning toolbox (http://mirlab.org) to compute the DTW measure.

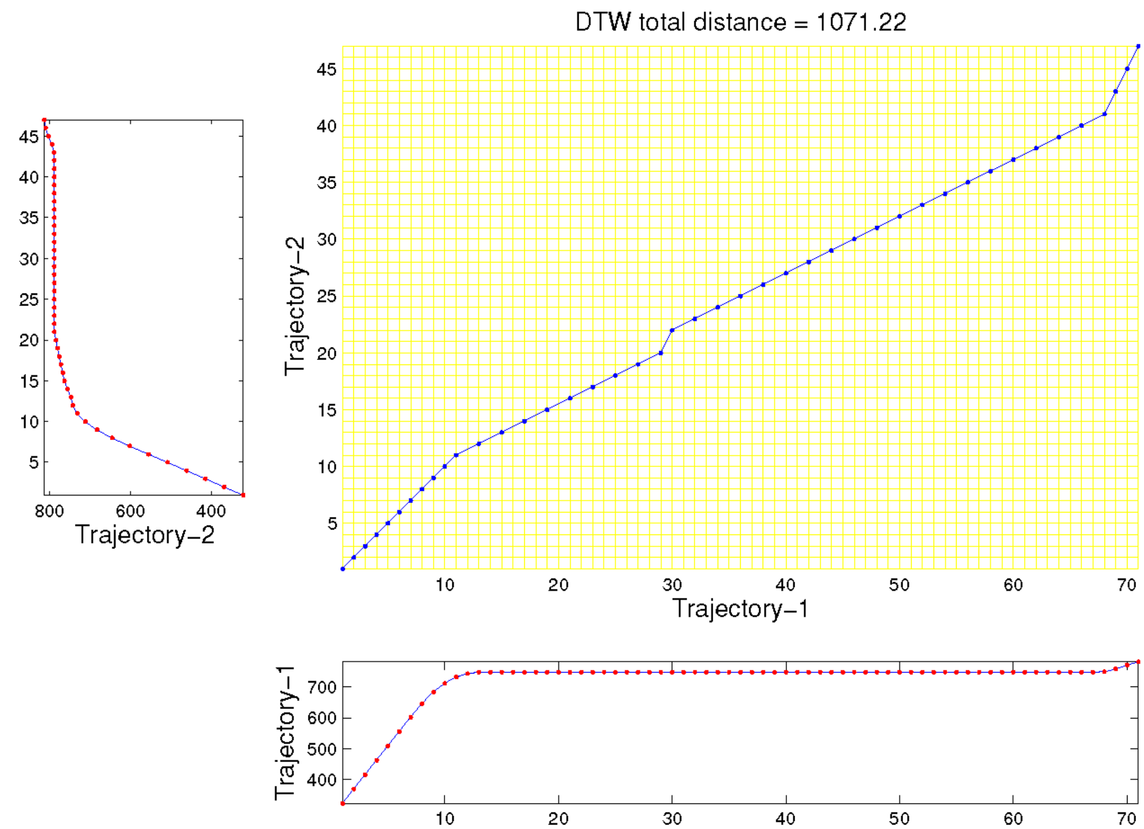


Figure 4. DTW optimal path computation for vehicle trajectories

4.2 Hierarchical clustering

Hierarchical clustering partitions the data in sequential steps and the number of clusters is not predefined. The clustering method yields a series of partitions that may include a single object or all the objects in the data. Further classification of hierarchical clustering leads to two subdivisions: agglomerative and divisive methods. This research follows agglomerative method that starts with all objects into a single cluster and ends with each object as a distinct cluster. At each stage of clustering agglomerative method fuses objects or groups which are most similar. The similarity measure dominates the patterns of partitioning. Finally, a binary hierarchical tree (linkage) is found. Clustering can be implemented with different specification of linkage. Single linkage (also known as nearest neighbor) defines the group distance as the distance between the closest pair of points of each group. Complete linkage (farthest neighbor) refers to the case where group distance is defined as the distance between the farthest pair of points of each group. Average linkage clustering defines group distance as the average distance between all pairs of points. The results reported in this research are obtained from average linkage hierarchical clustering. After obtaining the dendrogram (binary tree) one needs to decide where to cut the tree (in other words we need to determine the number of clusters). This research follows two techniques: (1) cutting at an height that yields desired number of clusters, (2) cutting the inconsistent links using inconsistency coefficients.

4.3 Consistency measures: Inconsistency coefficient

The comparison of the height of each link in a cluster tree with the heights of neighboring links below it in the tree can provide a direction to determining optimal number of cluster in a tree (Everitt et al., 2001; Gerbec et al., 2002). If two links (vertically spaced) have almost the same height, the cluster distinction between these two levels is not significant. These links are characterized by higher level of consistency. Obviously links with lower level of consistency (or inconsistency) are desired. Higher level of inconsistency indicates that the distance between the objects being joined is approximately the same as the distances between the objects they contain. The links in the dendrogram that have significantly different height compared with the height of the links below can be identified as inconsistent. Outputs from hierarchical clustering can be used to determine the the relative consistency of each link that is denoted as the inconsistency coefficient. The coefficient is a ratio of the height of the link to the average height of all links that lie below the link. High value of inconsistency coefficient indicates the joining of distinctly featured clusters. On the other hand, low value indicates that the clusters may be insignificant and possibility of merging is high.

4.4 Similarity measures: Cophenetic correlation coefficient

A property of a hierarchical cluster tree is that any two data objects must be connected to each other at some hierarchical levels of the tree. The cophenetic distance between any two objects is defined as the distance between the two clusters that contain these two objects. The distance between two clusters is measured as the height of the link in the tree. Intuitively, if the clustering is valid, the linking of objects in the cluster tree will be strongly correlated with the distances (DTW measures in our case) between objects in the original data. The cophenetic correlation coefficient measures this correlation between linkage distances and (dis)similarity distances (Farris, 1969; Lessig, 1972). A cophenetic coefficient closer to 1 indicates higher validity of the output from hierarchical clustering.

CHAPTER 5. DOMONSTRATION OF THE TECHNIQUE

The applicability of the proposed technique is demonstrated with the data from of a corridor with five signalized intersections, shown in Figure 5. The corridor islocated in the city of West Lafayette, Indiana in the U.S. The trip demand and signal settings are obtained from a previous study (Nichols and Bullock, 2001). The signal settings are based on Econolite controllers and designed for maximum congestion period. The network along with demand and signal settings is simulated using commercial traffic simulator VISSIM PTV America (2012) and trajectories are collected for 15 minutes period.

5.1 Comparing the performance of HC-DTW

To demonstrate the applicability and benefits of the HC-DTW technique, we compare the emissions estimates obtained from three different methods. The first method is the HC-DTW technique that uses the Link Driving Schedule (LDS) set obtained by clustering. Second method uses average speed of the link to estimate emissions which is the most commonly used technique in practice. The third method uses trajectories of all vehicles on a link as input for MOVES. The last technique provides the most accurate estimate of emissions from MOVES because the activity of each vehicle is considered as a separate link input into MOVES. For instance, if we have 1000 vehicles traversing on a link during the analysis period, we have to create 1000 instances of the link in the MOVES input file. Since each vehicle's activity is considered separately, different operating modes are considered by MOVES with higher accuracy. Hereafter we refer to

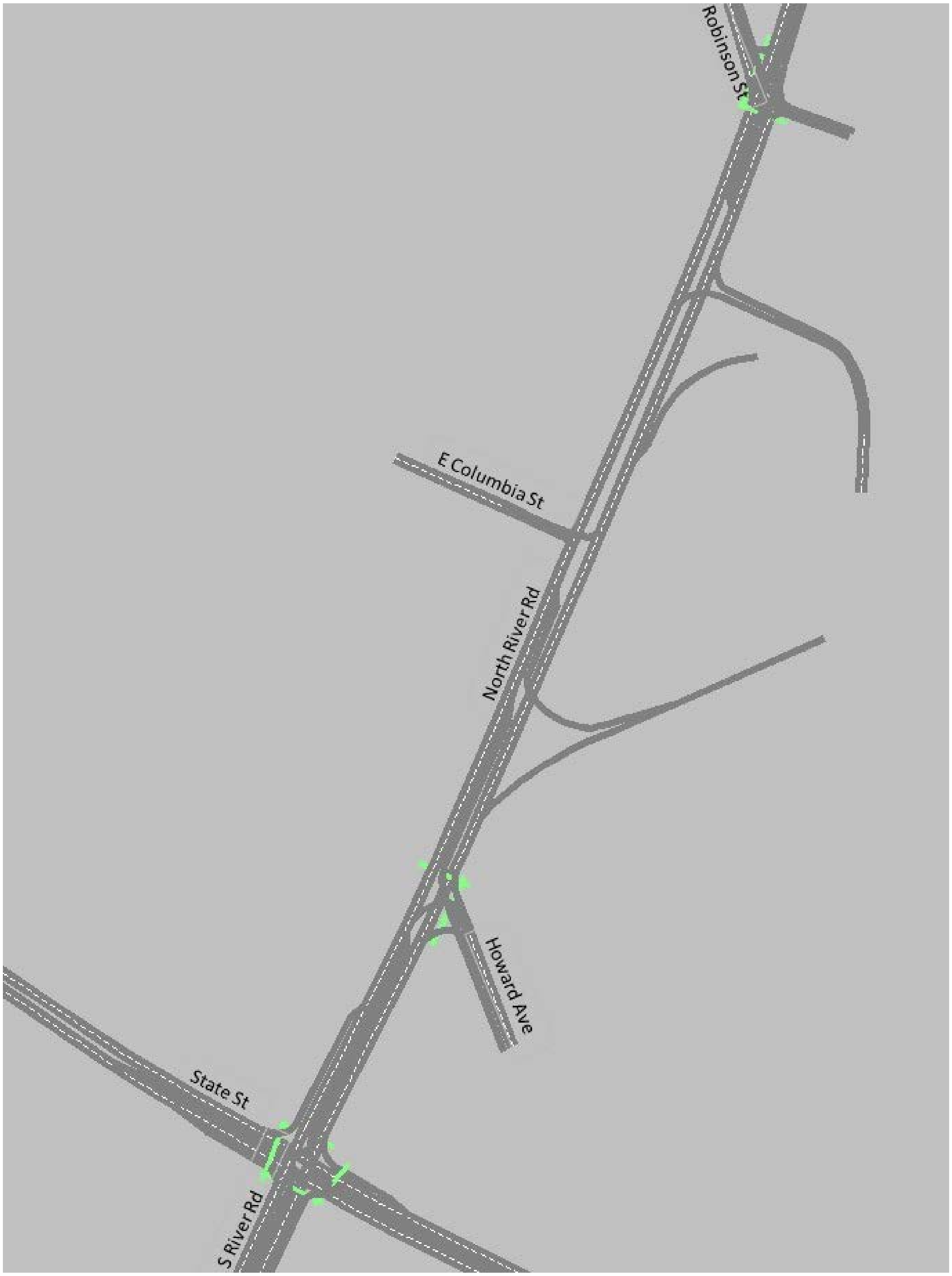


Figure 5. Test network: five intersection corridor

this approach as the exact method. The goal of this section is to demonstrate that the emissions obtained from our proposed approach is not significantly different from the those obtained from exact method.

We tested our technique with seven scenarios with the trajectories from the links of the signalized corridor. The test cases are constructed in such a way that variation in vehicular activities, resulting from varying congestion level or simply due to the existence of traffic signals, on road links can be represented. Although many possible combinations of vehicular activities exist, six types of activities are considered: (I) Cruise-Decelerate-Idle, (II) Idle-Accelerate, (III) Cruise, (IV) Cruise-decelerate-Accelerate, (V) Idle-Accelerate-Decelerate-Idle, and (VI) Idle. Table 1 shows the percentage of each type of vehicular activities for each case considered in this study.

Table 1. Description of test scenarios: vehicular activity proportions

	Vehicular Activity (%)					
Test Scenario	Cruise-Decelerate -Idle	Idle-Accelerate	Cruise	Cruise-Decelerate -Accelerate	Cruise-Accelerate-Decelerate -Idle	Idle
Case 1	31		21	48		
Case 2	40			55		5
Case 3	51	8	41			
Case 4		44		41		15
Case 5				46	42	12
Case 6	48	5			47	
Case 7		9		51	40	

For case 1, 39 vehicles traverse the road segment during the analysis. To compute the emissions using the exact method, we input 39 LDS (one LDS for each vehicle) in MOVES and compute the emissions for the link. Next, we apply the HC-DTW technique and we find three representative clusters. A prototype from each cluster is obtained to find the LDS. These LDSs are used as input for MOVES with corresponding traffic volumes. Likewise we estimate the emissions for all other cases. Further, we report the emissions estimated by inputting simply the average speed of the link for all cases (case 1 to case 7). Table 2 reports the emissions from these three approaches.

Further table 2 reports the percentage difference from the exact method for average speed and HC-DTW techniques. The differences for average speed method are found to be significantly higher compared with the HC-DTW technique. One intuitive cause is that the average speed method is not able to capture variations in speed profiles. For case 1, we observe 12 vehicles belong to a cluster where they do not have to stop on red phase of the signal or simply cruise through the link. The other two clusters (with 19 and 8 vehicles respectively) are characterized with acceleration, cruising, and then deceleration phase. Average speed for cluster-1 is about 22 mph and for the other two clusters the average speed is about 7.5 mph. The average speed method overestimates the speed for the link and uses a value of 13 mph and uses the corresponding driving cycle to estimate the emissions. This causes underestimation of emissions for that link. At lower speed (0-10 mph) the emissions are high, the average speed approach fails to capture that. Whereas by clustering we find representative speed profiles and emissions values are estimated with three LDS inputs that are close to exact method (table 2).

The difference values for PM_{10} is very high for cases 2, 3, and 7 with HC-DTW technique. However the difference values are still smaller compared with average-speed technique. The emissions estimate for PM_{10} are very small in terms of magnitude. It is possible to have accumulated error that leads to high percentage error although the absolute errors are small. HC-DTW provides one LDS for each cluster as a prototype of all vehicular activities in the cluster. Smaller difference in speed profiles between a member and the representative LDS can lead to higher degree of percentage difference as well.

Table 2. Comparison of estimates for different approaches using MOVES. **differences are from the exact method when each vehicle trajectory is an input

Case 1 (39 vehicles)					
Pollutants	Exact Method	HC-DTW	**diff. (%)	Avg speed	**diff. (%)
CO_2 (g/hr)	3607.66	3692.06	2.33	2905.75	19.42
CO (g/hr)	32.75	30.98	5.41	23.99	26.74
PM_{10}	0.1039	0.1059	1.91	0.0526	49.34
NO_x (g/hr)	2.879	3.035	5.40	2.164	24.82
Case 2 (62 vehicles)					
CO_2 (g/hr)	14884.44	13764.04	7.53	5939.55	60.10
CO (g/hr)	74.99	63.11	14.5	44.44	40.74
PM_{10}	0.27	0.22	18.88	0.10	63.94
NO_x (g/hr)	5.94	4.75	17.8	3.66	38.36
Case 3 (42 vehicles)					
CO_2 (g/hr)	4692.43	4603.01	1.91	4291.64	8.54
CO (g/hr)	35.34	34.00	3.77	29.77	15.75
PM_{10}	0.11	0.13	22.10	0.06	40.50
NO_x (g/hr)	3.42	3.23	5.41	2.24	34.39
Case 4 (34 vehicles)					
CO_2 (g/hr)	12469.52	11739.35	5.86	17401.29	39.55
CO (g/hr)	58.30	53.02	9.06	93.21	59.88
PM_{10}	0.22	0.19	13.45	0.22	2.38

Case 1 (39 vehicles)					
Pollutants	Exact Method	HC-DTW	**diff. (%)	Avg speed	**diff. (%)
NO_x (g/hr)	4.36	3.90	10.55	6.44	47.73
Case 5 (26 vehicles)					
CO_2 (g/hr)	11704.87	11135.34	4.87	16588.73	41.73
CO (g/hr)	52.60	49.71	5.50	88.62	68.47
PM_{10}	0.20	0.18	10.60	0.21	5.26
NO_x (g/hr)	3.88	3.62	6.72	5.97	53.97
Case 6 (36 vehicles)					
CO_2 (g/hr)	4282.95	3908.51	8.74	3738.09	12.72
CO (g/hr)	33.14	26.99	18.56	25.74	22.35
PM_{10}	0.11	0.10	2.05	0.06	47.84
NO_x (g/hr)	3.12	2.58	17.38	1.92	38.47
Case 7 (34 vehicles)					
CO_2 (g/hr)	3893.57	3845.91	1.22	3570.03	8.31
CO (g/hr)	30.07	30.00	0.22	24.90	17.19
PM_{10}	0.09	0.11	21.68	0.05	42.25
NO_x (g/hr)	2.86	2.74	4.09	1.89	34.04

Further we observe higher error percentage for CO emissions in case 2 and case 6. Both cases have higher percentage of cruise-decelerate-idle activity. It is possible that one single prototype of LDS cannot represent three activities: cruise, decelerate, idle. Higher number of clusters to represent cruise-decelerate, decelerate-idle, and idle separately can improve the results. We observe that the difference is higher as we increase the number of links (compare case 1 and case 2). This is mainly because of the accumulation of errors for each vehicle. This can be addressed by increasing the number of clusters when number of vehicles gets higher.

5.1.1 HC-DTW with small variation of congestion

Further, we tested the technique for a case where the variation of speed profiles is small. Three clusters with slight variation were chosen and all represent cruise-decelerate-idle activity. We found the difference with Exact method are 2% and 2.5% for HC-DTW and average-speed techniques respectively. This implies that the benefits of HC-DTW technique is marginal in cases where the vehicular activities do not vary much.

5.2 Computational time

We also observe significant reduction in running time when HC-DTW technique is applied. Table 3 shows the computational time for each cases. To conclude from our results, we do not have to compromise accuracy and obtain the solutions in much faster time. The running time is a function of the links in the network. For large networks with links in the order of thousands, the computational time can go up to days (EPA, 2012a). With the HTC-DTW technique it is possible to reduce the computational time. The analyst can define the error tolerance and the corresponding number of clusters to be used for large scale analysis. Further, it is possible to use real-world data as input to this algorithm and that will expedite the emissions estimation process for local transportation agencies while assessing air quality for transportation networks.

Table 3. Computational time comparison (conducted with 2.4 GHz, i5 4 core processor, 4GB RAM machine)

Estimation Method	No. of LDS	Computational Time (minute)
	Case 1	

Exact Method	39	33
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute

Case 2

Exact Method	62	39
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute

Case 3

Exact Method	42	34
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute

Case 4

Exact Method	34	25
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute

Case 5

Exact Method	26	18
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute

Case 6

Exact Method	36	29
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute
	Case 7	
Exact Method	34	25
HC-DTW	3	Less than a minute
Average-Speed	-	Less than a minute

5.3 Results with two signalized intersections

This section illustrates results obtained for seven links (signalized intersections) using the HC-DTW technique. The results report clusters of trajectories. It is straight forward to convert a trajectory to a LDS, therefore we do not present results for LDS objects. Table 4 summarizes the results for seven links from the test network. Figures 6, 7, 8, 9, and 10 show examples of the clusters obtained and exhibit the heterogeneous vehicular activities on signalized intersections. The intersection at south-west corner in figure 5 is intersection-1 and the rest are labeled accordingly. Further the cophenetic coefficients have higher values close to 1 indicating higher degree of validity of the results obtained from clustering.

Table 4. Sample results using HC-DTW technique

Link-ID	Total no. of vehicles	No. of clusters	Min. no. of elements	Max. no. of elements	Cophenetic coefficient
Link-1	195	8	19	58	0.83
Link-2	197	5	17	77	0.96
Link-3	400	9	17	116	0.76
Link-4	364	7	13	94	0.8

Link-5	181	5	12	47	0.78
Link-6	186	8	12	28	0.74
Link-7	249	8	15	62	0.86

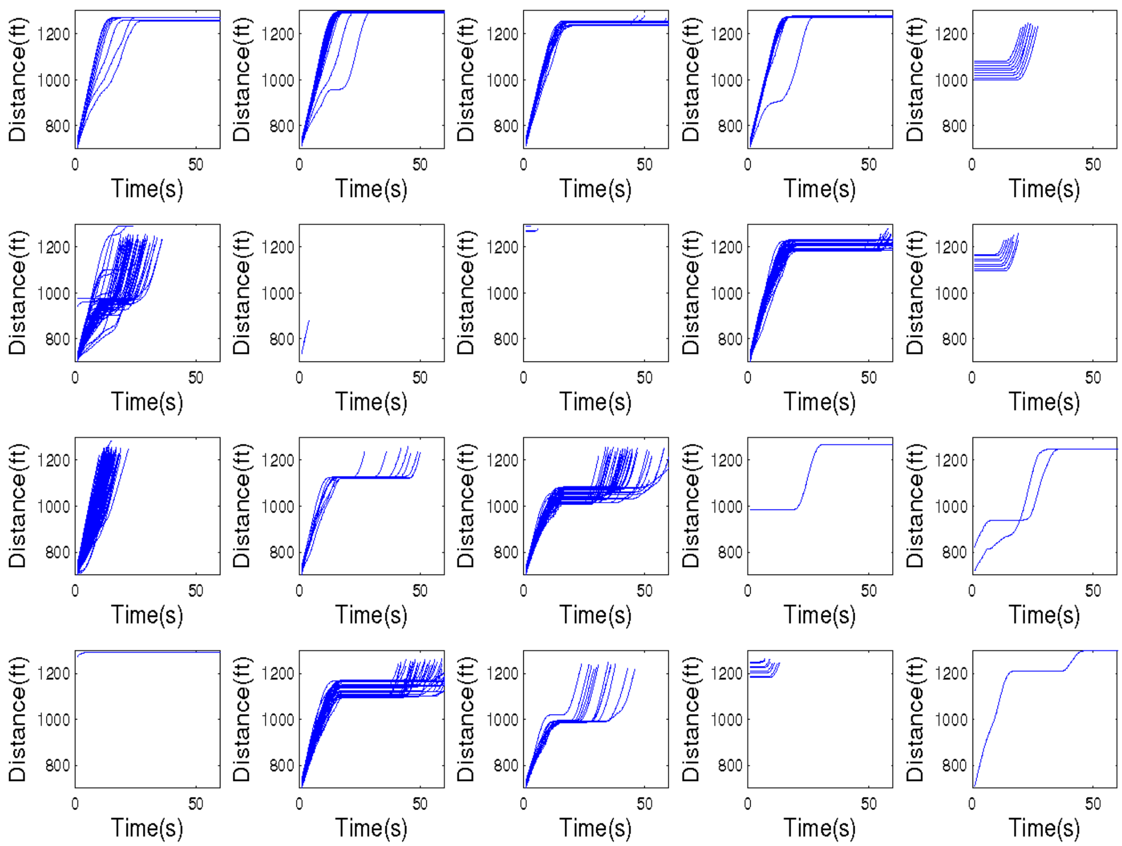


Figure 6. Clustering results for Link-3 in intersection-1

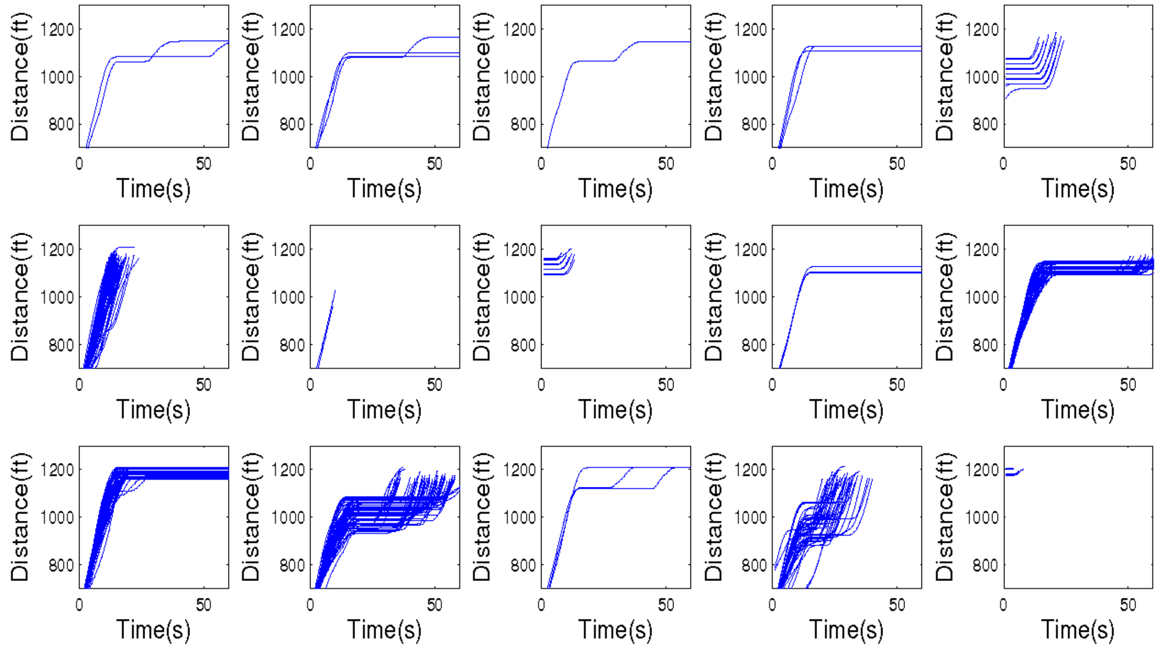


Figure 7. Clustering results for Link-4 in intersection-1

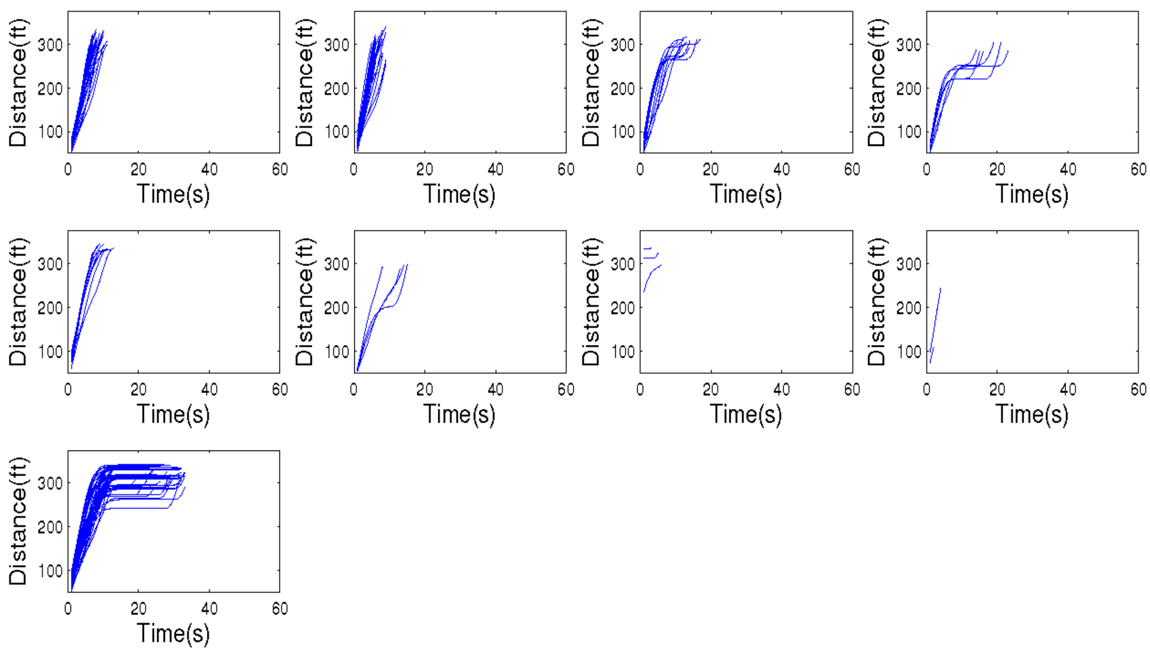


Figure 8. Clustering results for Link-1 in intersection-4

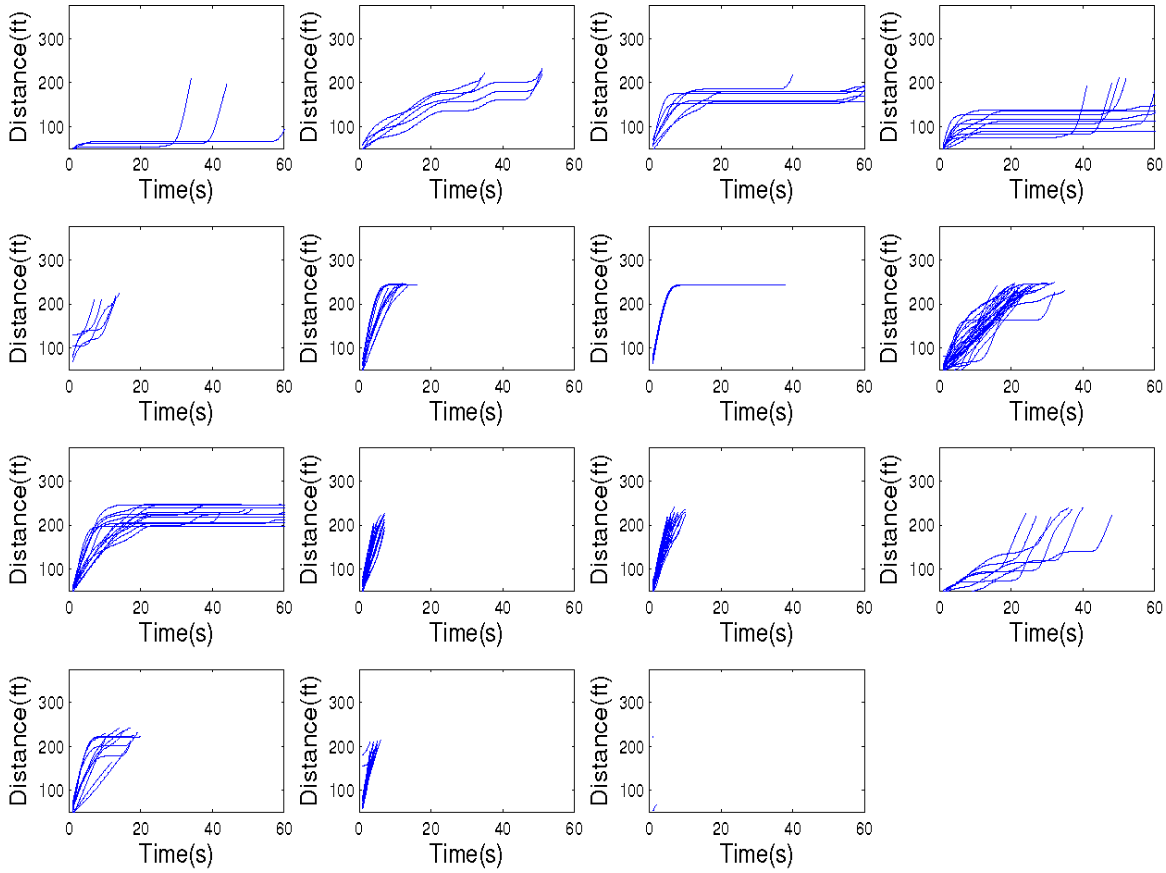


Figure 9. Clustering results for Link-2 in intersection-4

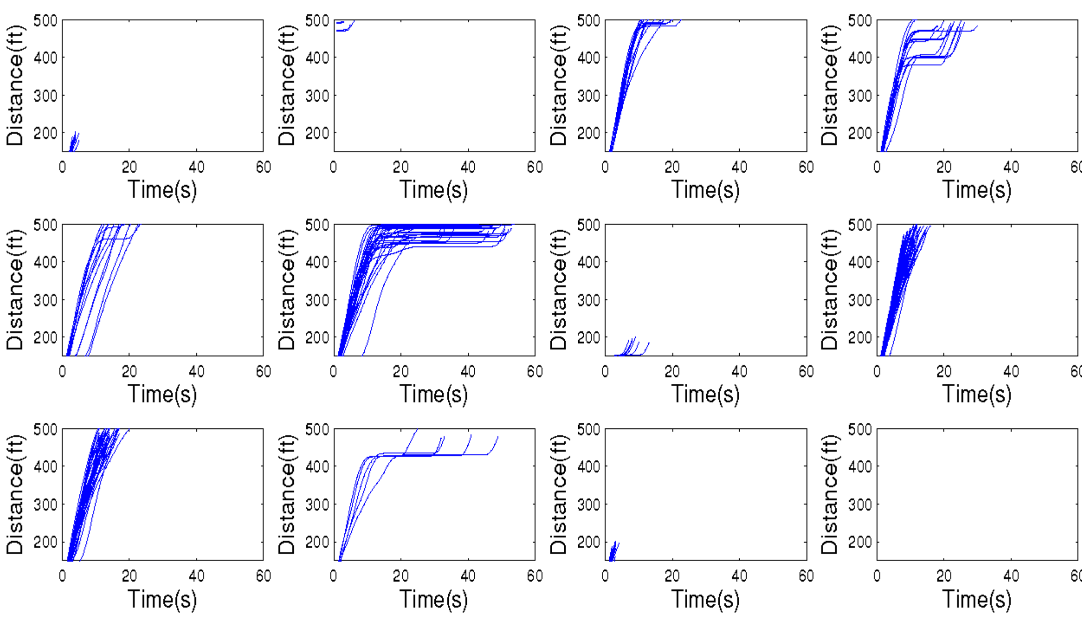


Figure 10. Clustering results for Link-3 in intersection-4

5.4 Summary of findings

The following conclusions can be made:

- The HC-DTWtechnique provides higher accuracy compared with averagespeed technique (table 2) as we see from our seven test cases.
- The error percentage for PM_{10} is found to very high in most cases. The results for $PM_{2.5}$ have similar trends and are not reported here for the sake of brevity.
- The number of links in a cluster affects the accuracy of the estimation. As the number of vehicle increases in a cluster, the degree of similarity (closeness) decreases. Although the average similarity remains same, some details are lost. Further, estimation with a very high volume traffic adds the accumulated error. This implies that the when the number of trajectories are reasonably high the analyst needs to carefully decide the number of clusters to include (section 4.2) as input to MOVES.
- Except PM_{10}, the error percentage ranges from 1% to 10% for all other pollutants. Case 2 and case 6 exhibit higher errors for NO_x.
- HC-DTW is most effective when the variation in vehicular activities on a link is high which is typically found during peak hour conditions and in urban areas (section 5.1.1). If the congestion level is low, and the vehicular activities are similar, average speed technique provides estimation with reasonable accuracy.

CHAPTER 6. CONCLUDING REMARKS

This research develops a novel technique to find link driving schedules from vehicular trajectories that can be used as input for emissions simulator MOVES. We developed a link driving schedule finding technique based on similarity based clustering that overcomes the general limitations of current approaches to estimate emissions. The hierarchical clustering based dynamic time warping (HC-DTW) technique is shown to have reasonably accurate results when compared with a exact method that treats each vehicle's speed profile as a distinct LDS. Analysis with a link shows that the difference in emissions are very small. In addition, the computational time is significantly smaller than the exact method. The major contributions are as follows:

(a) A similarity based clustering technique is proposed that finds the representative link driving schedules for MOVES input from second-by-second trajectory data.

(b) The hierarchical clustering based dynamic time warping (HC-DTW) technique does not compromise accuracy at a higher degree. This is particularly important for networks with highly varying congestion states. Traditional approaches such as average speed based techniques tends to under or over estimate the emissions, whereas HC-DTW technique is expected to have estimated emissions with very small difference with the exact method.

(c) The computational time decreases significantly because HC-DTW only uses cluster prototypes as the LDS for MOVES input. Since the number link-instances gets lower, the computation is faster.

The proposed technique can be used with any traffic simulation tool that provides second-by-second data. Moreover, real-world trajectory data can be easily processed and converted into LDS using our HC-DTW technique. The computational time also depends on the choice of number of clusters. However this will be decided by the analyst depending on the scope of the study.

For HC-DTW we observe that, the error level ranges from 2% to 18% in most cases. We observe the error level goes up with higher links in a cluster. This can be an issue if we estimate emissions for clusters with very high number of links. This can be addressed by increasing the number of clusters so that each cluster does not have a very high number of links. Also, more tests are required to have a stronger conclusion on the savings of computational time. Further, the technique is more promising to links with higher variability of vehicular activities. Nevertheless, the HC-DTW technique will significantly benefit the estimation of emissions using MOVES in terms of efficiency and accuracy and serve as a useful tool for the practitioners in context of air quality conformity analysis.

7. REFERENCES

Bapat, A. and H. O. Gao. (2010). Diesel Particulate Matter Number Emissions: Evaluation of Existing Modal Emission Modeling Approaches. *Journal of Transportation Engineering*, Vol. 136, No. 2, 2010, pp. 93-101.

Barth, M., F. An, T. Younglove, G. Scora, C. Levine, M. Ross, and T. Wenzel. (2000). Development of a comprehensive modal emissions model.

Barth, M., G. Scora, and T. Younglove. (2004). Modal Emissions Model for Heavy-Duty Diesel Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1880, pp. 10-20.

Berndt, D. J. and J. Cli_ord. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. In KDD workshop. Seattle, WA, Vol. 10, pp.359-370.

Boulter, P. G. and I. S. McCrae. (2007). The Links between Micro-Scale Traffic, Emission, and Air Pollution Models. Transport Research Laboratory, PPR 269.

Chamberlin, R., B. Swanson, E. Talbot, J. Dumont, and S. Pesci. (2011). Analysis of MOVES and CMEM for evaluating the emissions impact of an intersection control change. In Transportation Research Board 90th Annual Meeting, 2011.

Chu, S., E. Keogh, D. Hart, and M. Pazzani. (2002). Iterative Deepening Dynamic Time Warping for Time Series. Proceedings of the 2002 SIAM International Conference on Data Mining, pp. 195-212.

EPA. U., EPA Releases MOVES2010 Mobile Source Emissions Model: Questions and Answers. EPA-420-F-09-073, 2009.

EPA. U. S. E. P. A. Motor Vehicle Emission Simulator: User Guide for MOVES2010b. Vol. 37, No. EPA-420-B-12-001b, 2012a.

EPA. Using MOVES for Estimating State and Local Inventories of On-Road Greenhouse Gas Emissions and Energy Consumption. EPA-420-B-12-068, 2012b.

Everitt, B. S., S. Landau, M. Leese, and D. Stahl. (2001). Hierarchical clustering. Cluster Analysis, 5th Edition, pp. 71-110.

Farris, J. S. (1969). On the Cophenetic Correlation Coefficient. *Systematic Biology*, Vol. 18, No. 3, pp. 279-285.

Frey, H. C., K. Zhang, and N. M. Rouphail. (2010). Vehicle-Specific Emissions Modeling Based upon on-Road Measurements. *Environmental Science and Technology*, Vol. 44, No. 9, pp. 3594-3600.

Gerbec, D., S. Gasperic, I. Smon, and F. Gubina. (2002). An approach to customers daily load profile determination. Power Engineering Society Summer Meeting, 2002 IEEE, Vol. 1, pp. 587-591.

Hao, L., L. Shuang, Y. Zhuo, W. Heng, Y. J. Yang, and W. Xinhao. (2012). Framework for Integrating Traffic-Source Emission Estimates into Sustainability Analysis. In CICTP 2012, American Society of Civil Engineers, 2012, pp.3612-3619.

Huai, T., T. D. Durbin, T. Younglove, G. Scora, M. Barth, and J. M. Norbeck. (2005). Vehicle Specific Power Approach to Estimating On-Road NH3 Emissions from Light-Duty Vehicles. *Environmental Science and Technology*, Vol. 39, No. 24, 2005, pp. 9595-9600.

Jimenez-Palacios, J. L.. (1999). Understanding and quantifying Motor vehicle emissions with vehicle specific power and TILDAS remote sensing. Ph.D. thesis, MIT, Cambridge, MA.

Keogh, E. and C. A. Ratanamahatana. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, Vol. 7, No. 3, , pp.358-386.

Keogh, E. J. and M. J. Pazzani. (1999). Scaling up dynamic time warping to massive datasets. In Principles of Data Mining and Knowledge Discovery, Springer, pp. 1-11.

Kruskall, J. and M. Liberman. (1983). The symmetric time warping algorithm: From continuous to discrete. Time Warps, String Edits and Macro-molecules.

Lents, J. and N. Davis. International Vehicle Emission (IVE) Model Users Manual, 2004.

Lessig, V. P. (1972). Comparing Cluster Analyses with Cophenetic Correlation. *Journal of Marketing Research*, Vol. 9, No. 1, pp. 82-84.

Lin, J., C. Yi-Chang, S. Vallamsundar, and B. Song. (2011). Integration of MOVES and dynamic traffic assignment models for fine-grained transportation and air quality analyses. In Integrated and Sustainable Transportation System (FISTS), 2011 IEEE Forum on, 2011, pp. 176-181.

Liu, H. and M. Barth. (2012). Identifying the effect of vehicle operating history on vehicle running emissions. *Atmospheric Environment*, Vol. 59, No. 0, 2012, pp. 22-29.

Mahut, M. and M. Florian. (2010). Traffic Simulation with Dynameq. In Fundamentals of Traffic Simulation (J. Barcel, ed.), Springer New York, Vol. 145 of International Series in Operations Research and Management Science, pp. 323-361.

Nichols, A. and D. Bullock. Design guidelines for deploying closed loop systems. Joint Transportation Research Program, Vol. No. FHWA/IN/JTRP-2001/11, 2001.

PTV America, V. 5.4 User Manual. PTV Vision, 2012.

Sakoe, H. and S. Chiba. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, Vol. 26, No. 1, pp. 43-49.

Xie, Y., M. Chowdhury, P. Bhavsar, and Y. Zhou. (2012). An integrated modeling approach for facilitating emission estimations of alternative fueled vehicles. *Transportation Research Part D: Transport and Environment*, Vol. 17, No. 1, 2012, pp. 15-20.

Contacts

For more information:

Satish Ukkusuri

Purdue University

550 Stadium Mall Drive

7654942296

sukkusur@purdue.edu

NEXTRANS Center

Purdue University - Discovery Park

3000 Kent Ave.

West Lafayette, IN 47906

nextrans@purdue.edu

(765) 496-9724

www.purdue.edu/dp/nextrans