2021 IEEE Intelligent Transportation Systems Conference (ITSC)
Indianapolis, USA. September 19-21, 2021

An Automated Framework for Deriving Intersection Coordination Plans

Dhruv Mahajan[1] , Yashaswi Karnati[1] , Jeremy Dilmore[2], Anand Rangarajan[1] and Sanjay Ranka[1]

***Abstract*— The road network infrastructure (signal controllers and detectors) continuously generates data that can be transformed and used to evaluate the performance of signalized intersections. Current systems that focus on automatically converting the raw data into measures of effectiveness have proven extremely useful in alleviating intersection performance issues. However, these systems are not well suited for automatically generating recommendations or suggesting fixes as needed. In this work, we demonstrate the use of machine learning and data compression techniques to build a recommendation system. Specifically, we present an end to end solution for automatically generating intersection coordination plans.**

I. INTRODUCTION

Coordination involves synchronizing multiple intersections to enhance the operation of directional movements in a system [1]. Typically, the green times for a set of intersections are synchronized along the primary directions of movement using (speed-dependent) timing offsets to account for the travel time between intersections. Such coordination of intersections can result in improvements in the quality of traffic flow along an arterial or street [2]. The current practice for coordinating and re-timing intersections remains a largely manual process [1], [2] and is based on temporally limited data samples. In this paper, we present a recommendation system that generates intersection coordination plans using high resolution intersection controller logs collected over large time periods.

The problem of coordinating intersections can be decomposed into two parts, namely:

1) Identification of intersections that should be coordinated and the times of day and days of week for which they should be coordinated. Coordination can have a negative impact on the quality of service on the side streets and hence should only be deployed when needed.
2) Offset computation: Once the set of intersections has been identified, green time offsets need to be computed such that the flow of traffic in the primary direction of movement can be maximized.

Our focus in this paper is to design an automated solution for the first problem, i.e., identification of intersections on a corridor that are good candidates for coordination and the time periods they should be coordinated for. Traditionally, to determine the intersections that should be coordinated and the time periods for which intersections should be coordinated, extensive data is collected for intersections, including human recorded observations. This can be both time and effort intensive, especially for large cities or regions with tens of corridors and thousands of intersections. Our approach, in this paper, is to leverage high resolution controller logs that are routinely collected at many intersections. These controller logs have been used to compute various measures of effectiveness for signalized intersections. Automated Traffic Signal Performance Measures (ATSPM) systems (UDOT ATSPM [3] and its derivatives) also use these logs to automatically compute many intersection performance measures.

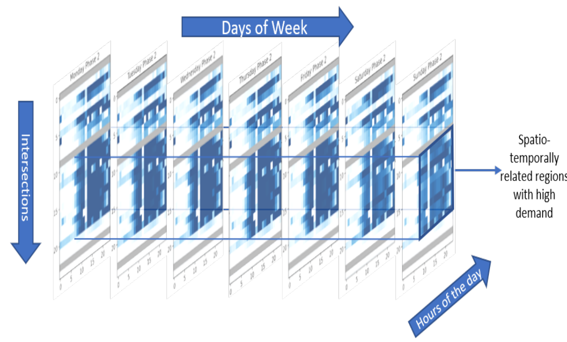

Fig. 1. A visual representation of the three dimensional nature of the data. Each plane in this figure represents a different day of the week. Intersections are plotted on the y-axis and hours of day are on the x-axis. The pixel color represents the intensity of the demand. Cuboids, like the one highlighted in the figure, represent spatiotemporally related regions of nearly homogeneous high demand in these data. We are looking for such regions as they represent subset of contigous corridors that should be coordinated during particular days of the week and hours of the day.

The problem of subdividing a single corridor into multiple spatially related sub-corridors with distinct traffic patterns or demand can be viewed as a 3-dimensional decomposition problem. The three dimensions in these data being time of the day (generally divided into hour size intervals or smaller), days of the week, and spatially correlated intersections. In essence, we are looking for spatially and temporally related regions with high demand in the 3-dimensional data shown in Figure 1. We need to look for similarity in traffic demand patterns across all three dimensions, while maintaining the spatial and temporal relations present in the data. This corresponds to cuboids with nearly homogeneous demand in Figure 1. And this will allow us to find sets of intersections (sub-corridors) that are good candidates for coordination because we expect them to have homogeneous performance.

*This work was supported by NSF CNS 1922782 and by the Florida Dept. of Transportation, District 5.

[1]Affiliated with the Dept. of Computer and Information Science and Engineering, Univ. of Florida, Gainesville, FL , USA {dhruvmhjn,yashaswikarnati,anandr}@ufl.edu, sranka@ufl.edu

[2]Affiliated with the Florida Department of Transportation, District 5, Deland, FL, USA jeremy.dilmore@dot.state.fl.us

2021 IEEE International Intelligent Transportation Systems Conference (ITSC) | 978-1-7281-9142-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITSC48978.2021.9564917

It will also allow us to deduce the times of day and days of week for which the coordination plans are required.

Our approach has two major steps:

1) Building accurate, descriptive models for performance measures of interest using time series modeling for all intersections under consideration. This is done to ensure that conclusions are not based on temporally local observations because they may be susceptible to outliers.
2) Use of these descriptive models (based on key MOEs) to cluster the intersections in a corridor with similar traffic patterns (or demand). This is done for all corridors under consideration. We use the results of clustering and spatial information within a corridor to automatically deduce contiguous regions in a corridor or sub-corridor where the intersections should be coordinated and the times of the day and days of week they should be coordinated for.

Our approach is scalable and can be applied to thousands of intersections. And can reduce the manual effort required during the process of corridor coordination significantly. This should allow for frequent changes in coordination times as the traffic evolves over time.

Our approach starts with identification of key MOEs, and the computation of those measures from the controller logs. Next, we use time series modeling to build representative models for each intersection and MOE under consideration. This is detailed in Section II. We use these models to then cluster intersections with similar traffic demand patterns. This is described in Section III. The spatial and temporal decomposition for each corridor is also described in Section III. Next, the results for a case study and a comparison with existing coordination plans is detailed in Section IV. Related work is presented in Section V. The conclusion and future direction for this work are presented in Section VI.

II. Intersection Level MOE Modeling

We employ descriptive modeling of the traffic demand patterns and then use the models to build a compact 3-dimensional representation of these data. Descriptive time series modeling is highly suited for capturing the periodic variations in traffic demand patterns. This also reduces the susceptibility to outliers. Using historical data to model intersection demand over a period of time would give us representative vectors for each intersection (I) and day (D) by capturing all the trends and variations in traffic demand patterns. We first describe our approach followed by results achieved on data collected from real intersections.

A. Descriptive Modeling

Based on input from practitioners and subject matter experts, we identified three key measures of effectiveness to model for this task. They are:

1) Demand-based split failures, computed using red and green occupancy ratios
2) Traffic volumes
3) Arrivals on red

These three MoEs were selected because they are useful in both coordinating intersections and diagnosing problems. Demand-based split failures and total volumes are both indicators of traffic demand patterns and hence these measures are modeled for each intersection. Modeling arrivals on red (number of arrivals when a phase is red) provides useful information to the practitioners, i.e. it is an indicator of correctness for the current (timing) offsets. Hence, we model all three measures. The first step is to aggregate the high resolution (10 Hz) data into minute-by-minute buckets. For split failures reported on a phase, a value of 1 is recorded during the minute the failure was reported. This is also true for minutes when more that one (demand based) split failure is reported. A 0 is recorded if there are no reported split failures for the phase. Using the computation in the previous paragraph (ROR/GOR), the split failures reported in coordinated corridors (when there is no demand) are ignored. Hence, we compute the duration of split failures in minutes and store that information as a time series vector. Similarly, we compute the traffic volumes and arrivals on red. The output of the first step is vectors with 24×60 (1,440) entries representing the behavior of a particular phase over the entire day. The next step in our methodology is to take these $1,440$-digit feature vectors and aggregate into one-hour bins. The dimensionality of the new vectors is 24, i.e., each vector has 24 entries representing one intersection for each the hour of the day. Note that the one-hour bin size was selected based on input from subject matter experts. However, various other bin sizes (15 min or 30 min) can also be explored. In our analysis, we have considered only the primary directions (phases 2 and 6) while creating these vectors and concatenated the vectors for phase 2 and phase 6 to create one vector for each MOE of interest.

The modeling step is critical because these models are effective in capturing time of the day and day of week variation in traffic demand patterns without being susceptible to temporal outliers. Hence, the modeling improves the current practice, which is based on temporally limited observations (one to two weeks of data).

We now describe the modeling step in more detail. The data that we aim to model are periodic in nature. Based on the previous research and input from subject matter experts, there are two periodic trends that need to be modeled: the daily and weekly periodicity of the data. Hence, we deploy Fourier series to model our data. The model can be described as follows:

$$y(t) = g(t) + s_1(t) + s_2(t) + e(t)$$

where,
$y(t)$ is the target variable, MOE in our case
$g(t)$ is the term that models growth over time
$s_1(t)$ is the term that models daily periodicity
$s_2(t)$ is the term that models weekly periodicity
$e(t)$ is the error term.

The periodic terms $s_1(t)$ and $s_2(t)$ can be further expanded as:

$$s_i(t) = \sum_{n=1}^{N}(a_n cos(\frac{2\pi nt}{P}) + (b_n sin(\frac{2\pi nt}{P}))$$

This modeling approach is adaptive to different periodic effects because of the reliance on Fourier series. The parameter P controls the periodicity of the time series under consideration. For the first periodic term, $s_1(t)$, the value of P is set to 24 to allow it to model daily trends. And for the second periodic term, $s_2(t)$, the value of P is set to 7 to allow it to model weekly trends. We build descriptive models for each intersection and measure separately.

Prophet™, an open source, state of the art, time series modeling and forecasting tool that was used to train these descriptive models. Periodic time series modeling in Prophet is also based on Fourier analysis, and hence it is well suited for our needs. To ensure the quality of the models, a test-train split was created in the dataset. Specifically, we used eight weeks of data to train the models and tested the models using the next four weeks of data. In all cases, the model accuracy is above 90%. This gives us a high degree of confidence that the models are capturing the periodic trends in the data. This approach allows us to build representative models for each intersection and MOE combination, thus enabling the comparisons across intersections that will be useful in the next steps.

B. Experimental results

Figure 2 is a visualization of the daily and weekly trends for all three measures of effectiveness that were selected. We expect that proximal intersections will report similar trends for split failures and volumes unless they are saturated. We can also observe that the volumes may be higher for certain days of the week, but the intersections are saturated (split failures) during the same times for all days of the week.

In the next section, we use the descriptive models built for demand-based split failures to cluster together intersections on the same corridor that are busy during the same times of day and days of week and then derive spatial decomposition of the corridors using this information.

III. Clustering and Spatiotemporal Corridor Decomposition

This section details the clustering of the intersections on the same corridor for different hours of day and days of the week using demand-based split failures. We pick this measure because it is a very good indicator of duration during which a intersection is unable to serve all the traffic demand. It is worth noting that other measures of traffic congestion can be used instead and the rest of the description does not depend on this choice.

Let there be multiple intersections $I = I_1, I_2, I_3, \ldots, I_n$, each of which is represented by its location (geographical coordinates). Let $M\ [I,\ D,\ H]$, D $\in$ [1,7] and H $\in$ [1,24] represent the modeled demand based split failures for a set of intersections for different days of the week and hours of the day. Although, we use hours as the discretization for decomposing the time of the day, one can use other interval sizes (15 minutes, 30 minutes) instead. These data can then be represented as a three dimensional cube, the three dimensions being intersections, days of the week and hours of the day.

For a set of intersections to be coordinated, they must be spatially proximal. However, the problem of finding spatially proximal regions of high demand in this 3D cube can be challenging. Hence, we first perform time of day clustering for all intersection-day (I-D) pairs. As a result of this, we obtain a single cluster identifier which represents the intersection demand patterns for the whole day. By clustering the I,D pairs, across the hour of day dimension we obtain a representative (cluster ID) for each pair and let this matrix be C[I,D].

Next, we deploy day of week (DoW) clustering to deduce which days of the week report similar demand patterns, and this effectively reduces our problem to a single dimension. This corresponds to looking at similarity between columns of C[I,D].

For each subset of columns found in the previous step, we look for nearby rows of the C matrix with similar behavior to derive subsets of similar intersections across the corridor with nearly homogeneous demand to generate candidates for coordination. A detailed explanation of each of these steps is now presented.

A. Hour of Day Clustering

As described above, we use the vector $M\ [i, d, *]$ to represent the performance of an intersection (i) for a single day (d). We propose the following steps to cluster the intersection-day (I-D) vectors and reduce our problem from a 3-dimensional to a 2-dimensional space:

1) First, we discretize the performance of each intersection, for each day, using demand-based split failures. Specifically, any observation that is $x\%$ of the peak demand for that intersection-day combination is classified as busy. We tried many values of x before settling on 55%. In this context, the peak demand is defined as mean of the three busiest hours under observation. These discrete vectors are represented as $B\ [I, D, H]$.
2) The next step is to deploy top-down hierarchical clustering combined with a custom distance measure to group similar intersection-day vectors together.
3) The last step is to compute the time of day cluster representative, $R[H]$, for each cluster. We compute the first eigenvector of set of intersection (I), day (D) vectors that get clustered together. This cluster representative, $R[H]$, summarizes the time of day traffic demand patterns for each cluster.

We use normalized Hamming Distance between vectors $B\ [i_1, d_1, *]$ for $B\ [i_2, d_2, *]$ for the purposes of clustering. This represents the number of hours for which the traffic congestion is different for two intersections. Hamming distance between two discrete vectors is defined as the number of positions the two arrays are different.

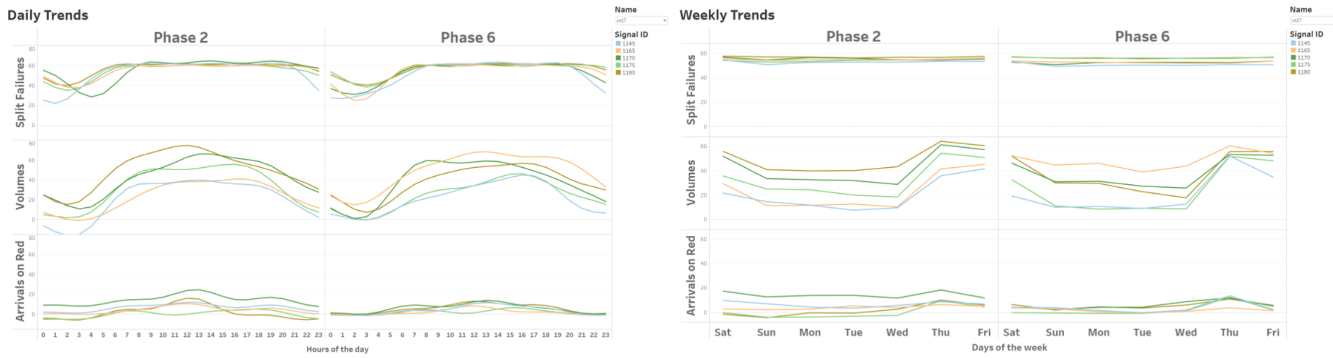

Fig. 2. A visualization of the daily and weekly trends captured by the models. Based on the daily trends for split failures and volumes, we can see that this set of intersections is busy from 5AM to end of the day. Arrivals on red are not correlated to the other two measures as they depend on the value of the timing offset. Although, not the focus of this paper, this can be useful in deriving the duration for each intersection where the offset can be improved. We can also see that most intersections report extended periods of saturation on all days.

This measure enables us to compare the times of day for which the two intersections are performing similarly. For example, if $B\ [i_1, d_1, *] = [1,1,0,0,0,0,1,1]$ and $B\ [i_2, d_2, *] = [1,1,1,1,0,0,1,1]$, the normalized distance will be $\frac{4}{8} = 0.5$. Figure 3 showcases how the clustering reduces the 3-dimensional problem to 2 dimensions.

Another important outcome of clustering is a representative vector for each cluster Y represented by vector $R[Y, H]$. $R[Y, H]$ corresponds to the hours when any intersection, day that is part of cluster Y is congested (and potentially has to be coordinated).

Next, we present the results of clustering for a sample corridor. $C\ [I,\ D]$, represent the time of day clustering results for all the intersections on a single corridor. Figure 3, part (B), is a pictorial representation of the matrix $C\ [I,\ D]$, i.e., the clusters that were discovered on a sample corridor. Each row of the image represents an intersection for 7 days. And the pixels are colored according to the respective cluster ID. The cluster ID's are ordered based on the increasing order of traffic demand. The rows that are grayed out represent bad data. We can see that two major regions of distinct intersection performance exist in this corridor. These are highlighted as S1 and S2.

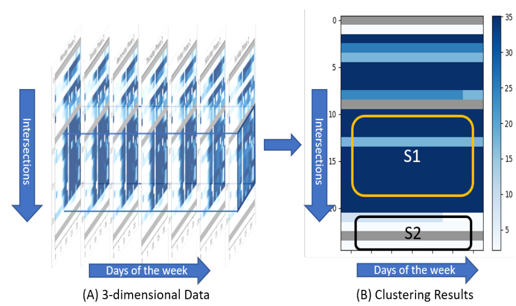

Fig. 3. Reduced representation of three dimensional data (A) into two dimensions. Additionally, (B) also presents the clustering results for a single corridor. Each row represents an intersection in the corridor and the columns represent the days of the week. Each cell is colored using the cluster ID. The gray color represents intersections in the corridor for which data is not available. S1 and S2 represent two regions of distinct intersection performance in this corridor. Our goal is to find these regions automatically.

B. Day of Week Clustering

Next, we use the results from the previous section and look for (not necessarily contiguous) days that are behaving similarly. Recall that $C\ [I,\ D]$ represents the time of day clustering results for all the intersections on a single corridor. This corresponds to the matrix presented in Figure 3. We cluster the seven, $C[I, *]$ vectors to find days of the week that report similar demand patterns.

The distance between $C[I,\ d1]$ and $C[I,\ d2]$ can then be defined by the number of dissimilar elements in the two vectors. However, a much better measure is to use the corresponding representative vectors for the two cluster ids this computation. Since each elements corresponds to a cluster id, this distance for two related cells is computed based on hamming distance between $R[C[I,\ d1], *]$ and $R[C[I,\ d2], *]$.

We can then use this measure to cluster all seven $C[I, *]$ vectors. This gives us the days of the week with similar demand patterns. Figure 4, parts (A) and (B), present a visualization of this process. For this example, we find only one cluster across all 7 days.

C. Spatial Decomposition of Corridors

This section describes the last step of this process, i.e., decomposing a corridor into spatially proximal sets of intersections (or sub-corridors) which are good candidates for coordination. Each corridor has constraints on the number of coordination plans that can be implemented on the corridor simultaneously. Too many sub-corridors have negative impact on the quality of progression of vehicles. Hence, we look for spatially proximal intersections with nearly homogeneous performance. This is done separately for each cluster of days founds in the previous step. The clustering is achieved by clustering adjacent rows of $C[I, D]$ (for only a subset of days) with similar behavior. Again, the distance between is computed using the hamming distance between corresponding R vectors as in the previous step. In addition to the above approach, the clustering approach allows for a small number of non-conforming intersections

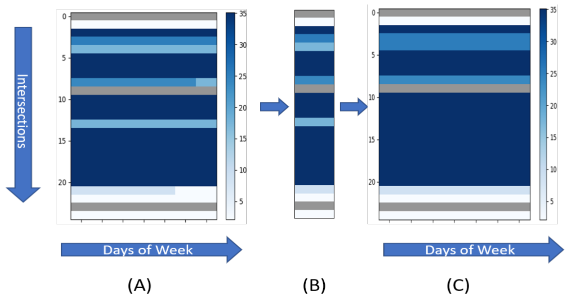

Fig. 4. The results of the algorithm used to detect spatially proximal, nearly homogeneous regions within a corridor such that these intersections can be coordinated together. (A) represents the results of the time of day clustering, (B) represents the results of the day of week clustering, and (C) depicts the regions detected by the spatial decomposition algorithm.

within a cluster of similarly behaving intersections. This is achieved by a two phase clustering where only similar adjacent intersections are clustered together. This is followed by merging two clusters with similar behavior if they are separated by a single intersection whose traffic congestion behavior is lower than the two adjacent clusters on either side that should be merged. This process is depicted by Figure 4, part (B) and part (C).

This concludes the automated approach for deducing coordination plans for a set of intersections on the same corridor and the times of day and days of week for which they should be coordinated. The next section presents a case study which showcases the results of this approach.

IV. Results and Case Studies

The techniques described in the previous sections were applied to approximately 200 intersections spread across 10 key corridors in Orlando, Florida. This section presents a case study for Lake Mary Boulevard which showcases our results. For the year 2019, we compare our results with coordination plan used during that year.

The results presented in Figures 5 and 6 were based on descriptive models built using data from 2019. Figure 5 details the clustering and spatial decomposition results for the corridor under consideration. We can see that one major (and one minor) region of distinct performance are present on this corridor. These are marked as A and B. The intersections corresponding to these regions are highlighted on the map. This figure also presents the demand on both the major and minor phases for comparison. It should be noted that our clustering is based on the major phase demand only. The minor phase demand can be useful in further sub-dividing these regions and hence is presented with our results. These results show that for subcorridor represented by B, the signals should be coordinated from 7 AM to 8 PM (this is shorter than the currently used times of 6 AM to 9 PM). For A, no coordination is necessary.

The next figure, Figure 6, presents a comparison of our results with current coordination plans can be improved using our recommendation. Specifically, the coordination time for one of the sub-corridors can be reduced. Choosing appropriate periods for coordinating signals on a corridor can have positive impact on the quality of service on the side streets. We also present the clusters from which these coordination plans were derived.

V. Related Work

Traffic engineers often face the challenge of quantifying the performance of signalized intersections and effectively troubleshooting day-to-day operational issues. In many cases, problems that are particularly difficult to troubleshoot are the ones that reoccur only at specific time of day or because of exclusive traffic patterns. A practitioner has to generate and analyze performance measures for each direction of movement, and for every intersection manually to understand and troubleshoot thee problems. There have been past attempts to address these shortcomings. Specifically, data analytics techniques have been previously applied to traffic flows and we now present a few relevant studies. Howell Li *et al.* [4] present a heuristic based on system wide split failure identification and evaluation. By using this heuristic, they demonstrated performance improvements for specific corridors. The paper titled *"Graphical Performance Measures for Practitioners to Triage Split Failure Trouble Calls"* [5], demonstrates the use of graphical performance measures based on detector occupancy ratios to verify reported split failures and other shortcomings in signal timing that are often reported to traffic engineers by the public. Wemegah *et al.* [6] present techniques for management of large signalized intersection datasets with the aim of analyzing traffic volumes and congestion, addressing all the steps in the analytics pipeline, namely, data acquisition, data storage, data cleaning, data analysis and visualization. Huang *et al.* [7] propose a set of new, derived MOE's that are designed to measure health, demand and control problems on signalized intersections. The newly proposed MOE's are based on approach volume and platooning data derived from ATSPM [3] systems.

VI. Conclusion

In this work, we presented an automated framework for deducing intersection coordination plans. To this end, we deployed time series modeling to build descriptive models for all the performance measures under consideration. Next, these models are used to cluster intersections with similar traffic demand patterns. Furthermore, we spatially decomposed the corridors into regions with near homogeneous traffic demand, generating sets of intersections that are good candidates for coordination. Lastly, we presented a case study to showcase our recommendations and compared these recommendations with the currently deployed coordination plans. This automated approach can be deployed in two ways. First, it can serve as a decision support system for traffic engineers and generate sets of intersections that are good candidates for coordination. Second, it can serve as a prior step for an automated signal re-timing solution.

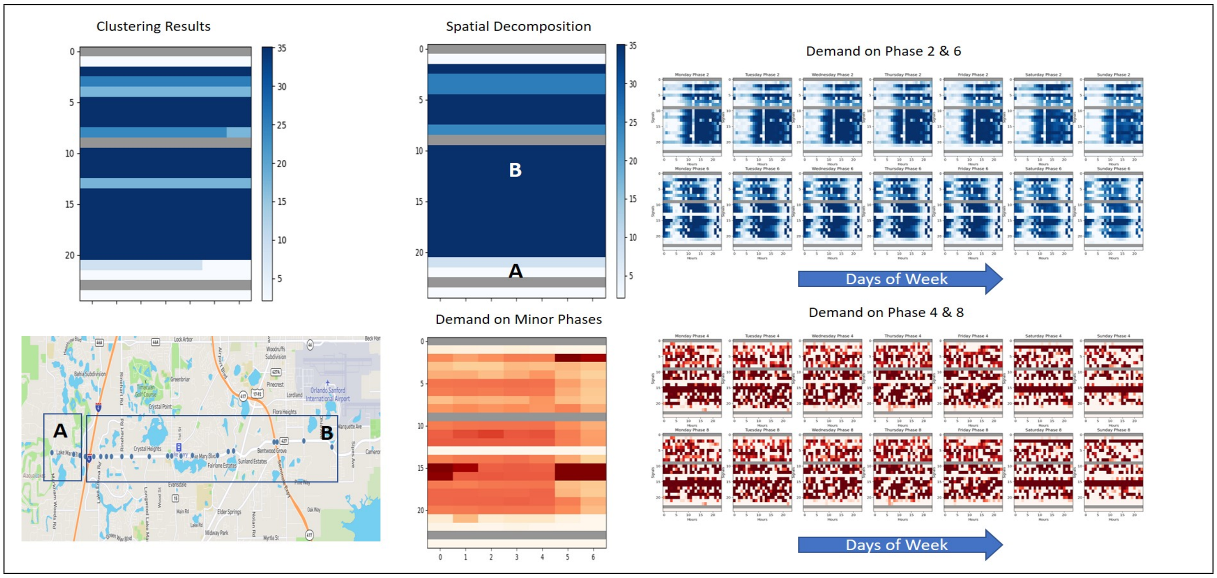

Fig. 5. The clustering and spatial decomposition results for a single corridor (Lake Mary Boulevard) using data from the year 2019. The two major sub-corridors are highlighted on the map. The split failures for the major (2,6) and minor (4,8) phases have also been plotted on the left. It should be noted that our clustering is based on the major phase split failures only.

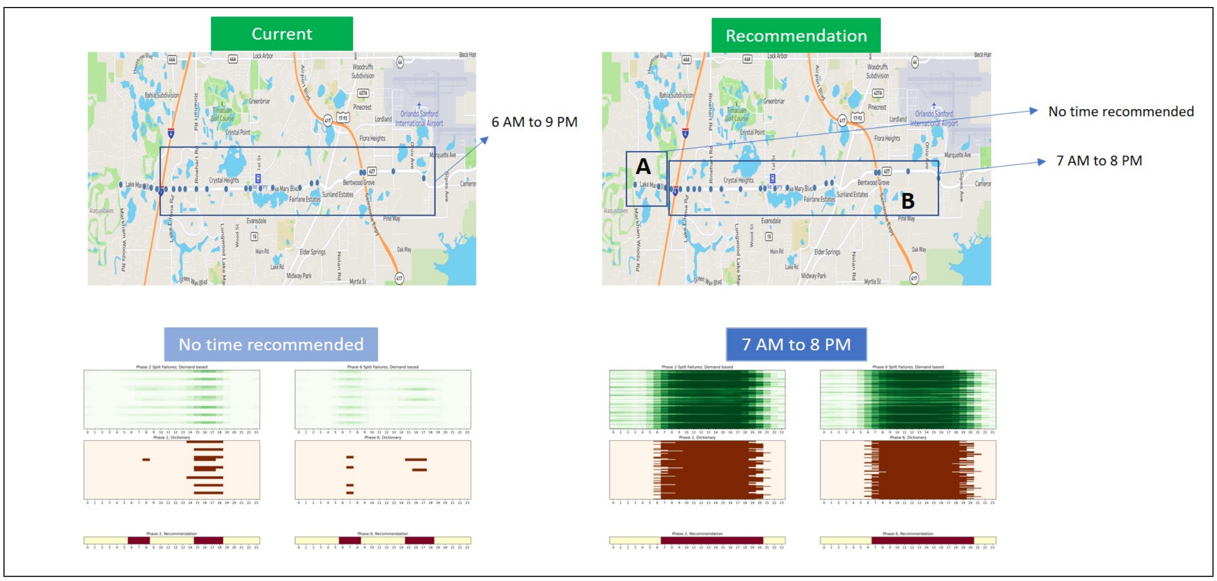

Fig. 6. We present the recommended coordination plan based on the data collected for Lake Mary Boulevard for the year 2019. Intersections in sub-corridor (B) are currently coordinated between 6 AM and 9 PM. Whereas the intersections in sub-corridor (A) are not coordinated for most times of the day. Thus, our recommendation can improve the existing coordination times.

REFERENCES

[1] F. H. A. US Department of Transportation, "Traffic signal timing manual," https://ops.fhwa.dot.gov/publications/fhwahop08024/index.htm, 06 2008, (Accessed on 2/10/2020).

[2] C. Day, T. Brennan, H. Premachandra, A. Hainen, S. Remias, J. Sturdevant, G. Richards, J. Wasson, and D. Bullock, "Quantifying benefits of traffic signal retiming," *Transportation Research Record*, p. pp, 01 2010.

[3] UDOT, "UDOT automated traffic signal performance measures - automated traffic signal performance metrics," https://udottraffic.utah.gov/atspm/, 2017, (Accessed on 2/26/2020).

[4] H. Li, L. M. Richardson, C. Day, J. Howard, and D. Bullock, "Scalable dashboard for identifying split failures and heuristic for reallocating split times," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2620, pp. 83–95, 01 2017.

[5] R. Freije, A. Hainen, A. Stevens, H. Li, W. Smith, H. Summers, C. Day, J. Sturdevant, and D. Bullock, "Graphical performance measures for practitioners to triage split failure trouble calls," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2439, pp. 27–40, 12 2014.

[6] T. D. Wemegah and S. Zhu, "Big data challenges in transportation: A case study of traffic volume count from massive radio frequency identification(rfid) data," in *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, Oct 2017, pp. 58–63.

[7] T. Huang, S. Poddar, C. Aguilar, A. Sharma, E. Smaglik, S. Kothuri, and P. Koonce, "Building intelligence in automated traffic signal performance measures with advanced data analytics," *Transportation Research Record*, vol. 2672, no. 18, pp. 154–166, 2018. [Online]. Available: https://doi.org/10.1177/0361198118791380