Discovering EV Recharging Patterns through an Automated Analytical Workflow

Preprint · August 2020

DOI: 10.13140/RG.2.2.18976.40966

3 authors:

Rene Richard
National Research Clouncil

Hung Cao
University of New Brunswick

M. Wachowicz
University of New Brunswick

Discovering EV Recharging Patterns through an Automated Analytical Workflow

René Richard[†*], Hung Cao[*], Monica Wachowicz[*]
[*] *People in Motion Lab, University of New Brunswick, Canada*
[†] *National Research Council of Canada, Digital Technologies Research Centre, Fredericton, Canada*
{rene.richard, hcao3, monicaw}@unb.ca

***Abstract*—The vision for smart cities is to provide a core infrastructure that enables a good quality of life for their citizens and the sustainable management of natural resources. Towards this vision, supporting the adoption of Electric Vehicles (EV) contributes to improved air quality, sustainable mobility, and utility distribution. Fostering EV adoption contends with concerns typically centered on vehicle range and costs. An understanding of EV charging patterns is therefore crucial for optimizing charging infrastructure placement and managing operational costs. Towards this end, this paper proposes an automated analytical workflow to gain insight from a large volume of real operational data from EV charging stations. The research goal is to establish a mechanism to descriptively analyse the EV charging data and to thoroughly diagnose whether low-demand charging station groupings can effectively be identified using spatio-temporal features and hierarchical clustering. Preliminary results suggest agglomerative clustering is effective at grouping similar charging stations together when considering spatial and temporal features of recharge events.**

***Index Terms*—agglomerative hierarchical clustering, EV adoption, charging infrastructure patterns, automated machine learning flow**

I. Introduction

Globally, national and local government commitments to electrify the transport sector will have an impact on smart cities. In Canada, 24% of GHG emissions come from cars. New Brunswick has the third highest per-capita GHG emissions in Canada and thirty percent of these emissions are attributable to the transportation sector [1]. By decoupling vehicles from the immediate consumption of fossil fuels, options for supporting mobility from renewable resources increase.

A significant increase in EV adoption will require adequate public charging infrastructure. One commonly stated barrier to widespread EV adoption is driver range anxiety [1], [2]. While increasing the availability of public charging infrastructure can appease some concerns, charging infrastructure operators do not want to invest in charging stations because they tend to be less profitable in early EV adoption contexts. Additionally, changing population demographics in some regions can pose further challenges to EV uptake. Enthusiasm around vehicle electrification has been associated with a youthful user base and financial incentives, which some regions struggle to provide [3].

Broader EV adoption will require a public recharging network to serve a population with different living and parking situations (e.g. multi-tenant dwellings) [4]. However, an understanding of EV usage patterns is also crucial to foster adoption while managing costs and optimizing placement of charging infrastructure. Although there exists a rich literature on different algorithms developed to find EV usage patterns, lacking is the use of real-world EV charging events from public infrastructure. Analyses are often based on assumptions that EV traveling profiles are similar to that of conventional internal combustion engine vehicles.

By using real-world EV charging event data, related demand characteristics can provide a better assessment of increased EV adoption and the potential impact on energy demand. The main challenge is to develop an analytical workflow that allows the creation of a containerized assembly, which fully automates analytical tasks and facilitates the reproducibility, sharing and distribution of the computational environment. Towards this end, this paper proposes an end-to-end analytical workflow to facilitate charging pattern analysis from an early EV adoption context in Atlantic Canada. In total, nine automated tasks are developed and a case study using real-world charging event data from station operators in Atlantic Canada is used for the implementation.

The main contributions of this paper are:

- Automated analytical tasks for discovering EV charging patterns are still nonexistent despite the need for analytical workflows capable of handling the expected volume and complexity of real-world EV charging events. The pioneering research work on the analytical workflow proposed in this paper is a first step towards this direction.
- The empirical results of this research work contribute to advance scientific knowledge on the current EV adoption patterns that play a role in planning and extending charging infrastructure design for electric vehicles.

The rest of the paper is organized as follows. In Section II, we explore related works. Section III describes the nine automated tasks developed for our proposed analytical workflow and Section IV provides a detailed description of real-world EV charging event data and computational environment used for implementation. In Section V, we discuss the results. Finally, Section VI concludes and indicates future research work.

II. Related Work

A significant proportion of the work studying EV charging patterns uses data from sources other than real-world EV charging events in order to assess the impact of broad EV adoption on distribution networks. This is apparent in a recent review paper by Hardmana et al. [5] on consumer preferences with regards to EV charging infrastructure, which lists studies that employed surveys, interviews, modeling and vehicle GPS data in addition to a small number of studies using EV charging equipment information. By using real-world charging event data, related demand characteristics can provide a better assessment of increased EV adoption and the potential impact on energy demand.

A more complete picture of EV infrastructure usage patterns can be formulated by combining data from different sources. However, scarcity in complimentary data sources is also a concern. Ashkrof et al. [2] have studied EV user travel behaviors in the Netherlands and point out the main limitations of their work was related to the low number of Battery Electric Vehicle (BEV) driver participants. In order to compensate for this, Hybrid Electric Vehicle (HEV) and Plug-in Hybrid Electric Vehicle (PHEV) drivers were added to the study. Despite the limitation in targeting BEV owners uniquely, the authors found that route attributes such as travel time, charging infrastructure characteristics en route to and at the destination of travel in addition to charging wait times and State-of-Charge (SOC) influences route selection and charging behavior considerably. In their study, the authors note that SOC and lack of charging opportunities are the main concerns of EV drivers.

The classification and modeling of energy usage behavior is core to improving emerging applications and services. Enhancements in services allow for smoothing of frequent peaks and imbalances. In the energy domain, clustering is commonly used to group similar consumers, predict future energy demand or detect outliers.

Our use of real operational data from EV charging stations advances our understanding of EV charging behavior. To the best of our knowledge, no other work has implemented an EV charging analysis workflow that facilitates automated, parallel, spatio-temporal analysis in a containerized environment. This work aims to answer the question of whether agglomerative clustering can effectively identify low-energy demand clusters by grouping under-utilized (and normally utilized) charging stations together using spatio-temporal features.

III. Methodology

The overall goal of building an analytical workflow is to explore patterns that develop around EV charging infrastructure usage. Therefore, the proposed workflow supports automated tasks that are essential in providing new insights on whether low-demand charging stations can effectively be identified to ultimately assess their impact on EV adoption (i.e. Descriptive analytical results). These automated tasks are also important to diagnose the cause of phenomena observed at charging stations based on clustering of real-world charging events (Diagnostic analytical results). Figure 1 illustrates the data flow and processing tasks of the proposed end-to-end analytical workflow which is designed to operate in an fully automated manner and facilitates the reproducibility, sharing, and distribution of the computational environment.

In total, nine automated analytical tasks are designed to analyse a massive amount of real-world EV charging data in a fast parallel processing environment using different techniques to track, manage, and compare the analytical results.

Each analytical task can be described as one of the following:

- **Data Cleaning Task:** This task receives the raw data input from the public charging stations. The raw data can be defined as a set of files $\mathbb{R} = \{R_1, R_2, ..., R_n\}$ where each file $R_i \ni [r_1, ..., r_m]$ data rows in which each data row r_j owns k attributes. The goal of this data cleaning task is to clean each raw data file in the set $\mathbb{R}$. A core function is developed to guarantee the data quality and produce a set of cleaned files $\mathbb{C} = \{C_1, C_2, ..., C_n\}$ by eliminating errors, inconsistencies, duplicated and redundant data rows, and handling missing data.
- **Data Integration Task:** This task is known as a practice of consolidating data from various data files into a single dataset. A variety of files from the cleaned set $\mathbb{C}$ in the previous task can be used as the input for this task. The output of this task is a unique file $\mathbb{I}$ that merged all attributes from set $\mathbb{C}$ into one big table.
- **Data Fusion Task:** Different from the data integration task, a data fusion task usually involves combining multiple data sources followed by reduction or replacement for the purpose of better inference. In our proposed analytical workflow, multiple integrated data files $\mathbb{I}$ can be pushed into the data fusion task and produce more consistent, accurate, and useful data files $\mathbb{F}$ that serve a more narrow set of application workloads.
- **Data Contextualization Task:** The aim of this task is to enrich the fused data files $\mathbb{F}$ step by step by adding new attributes to each data row according to a specific context. This task is defined by a contextualization function that can produce a set of new data rows $P_i \in \mathbb{P}$ using contextualization parameters $\langle \Psi_1, \Psi_2, ... \rangle$ to add new attributes to the fused data rows $F_i \in \mathbb{F}$.

$$\begin{cases} \forall F_i \in (F_1, F_2, \dots, F_n) : F_i = (f_1, ..., f_m) \\ \mathbb{F} = (F_1, \dots, F_n) \xrightarrow{\langle \Psi_1, \Psi_2, \dots \rangle} \mathbb{P} = (P_1, \dots, P_n) \\ \forall P_i \in (P_1, \dots, P_n) : P_i = (p_1, ..., p_m, Context_1, Context_2, \dots) \end{cases} \quad (1)$$

  It is crucial in transforming fused data rows generated by EV charging events into semantically enriched data that are needed as an input to the next analytical tasks.
- **Data Descriptive Task:** To gain an overall understanding from the contextualized data in the previous task, this task performs several descriptive statistical functions. They include frequency measurement, central tendency measurement, dispersion or variation measurement, and position measurement.

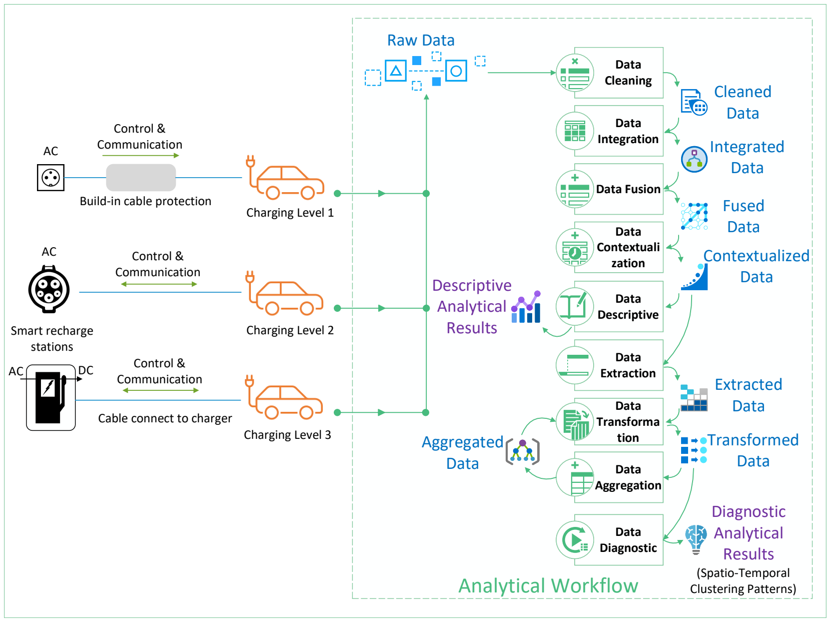


Fig. 1: The proposed automated end-to-end analytical workflow.

- **Data Extraction Task:** This task is defined by an extraction function that can produce a subset of data rows $\mathbb{E}$ that are extracted from a set of contextualized data rows $\mathbb{P}$ using extraction (filtering) parameters $\langle \Omega_1, \Omega_2, ...\rangle$ executed on a selected attribute (or a set of selected attributes) of a set of contextualized data rows $\mathbb{P}$.

$$\begin{cases} \forall P_i \in (P_1, \dots, P_n) : P_i = (p_1, ..., p_m, Context_1, Context_2, \dots) \\ \mathbb{P} = (P_1, \dots, P_n) \xrightarrow[\text{on attributes } (e_i)]{\langle \Omega_1, \Omega_2, \dots \rangle} \mathbb{E} = (E_1, \dots, E_n) \\ \forall E_i \in (E_1, \dots, E_n) : E_i = (att_1, att_2, \dots), \quad \forall att \subset \{e_1, \dots, e_m\} \end{cases} \tag{2}$$

- **Data Transformation Task:** This task is defined by a transformation function that performs transformation operations using parameters $\langle \Upsilon_1, \Upsilon_2, ...\rangle$ executed on a selected attribute (or a set of selected attributes) of a set of extracted data rows $\mathbb{E}$ or aggregated data rows $\mathbb{A}$ to produce a set of new data rows $\mathbb{T}$.

$$\begin{cases} (E_1, \dots, E_n) : E_i = (att_1, att_2, \dots) \\ \mathbb{E} = (E_1, \dots, E_n) \vee \mathbb{A} = (A_1, \dots, A_n) \xrightarrow{\langle \Upsilon_1, \Upsilon_2, \dots \rangle} \mathbb{T} = (T_1, \dots, T_n) \\ \forall T_i \in (T_1, \dots, T_n) : T_i = (Trans\_value_1, Trans\_value_2, \dots) \end{cases} \tag{3}$$

- **Data Aggregation Task:** Aggregation is a mathematical operation (e.g. sum, average, count, minimum) that takes multiple attributes of many rows and returns a single value. This task is defined by aggregation parameters $\langle \Phi_1, \Phi_1, ...\rangle$ executed on a selected attribute (or a set of selected attributes) of a set of transformed data rows $\mathbb{T}$ to produce a set of new data rows $\mathbb{A}$.

$$\begin{cases} \forall T_i \in (T_1, \dots, T_n) : T_i = (Trans\_value_1, Trans\_value_2, \dots) \\ \mathbb{T} = (T_1, \dots, T_n) \xrightarrow[\text{on attribute } Trans\_value_i]{\langle \Phi_1, \Phi_1, \dots \rangle} \mathbb{A} = (A_1, \dots, A_m) \\ \forall A_j \in (A_1, \dots, A_m) : A_j = (Agg\_value_1, Agg\_value_2, \dots) \end{cases} \tag{4}$$

- **Data Diagnostic Task:** The aim of this task is to find the patterns from transformed data using a hierarchical agglomerative clustering algorithm [6]. The raw recharge events are transformed into a per-minute kWh energy delivery format for each station recharge event, which also includes station latitude and longitude coordinates. Then, the PCA technique is used to reduce the dimensions of the transformed data, since the number of spatial and temporal features is very high. Finally, an agglomerative clustering model is utilized to fit this data. The clustering was performed on weekly, monthly and seasonal data partitions to provide results for different time windows. A priori unknown schemes inherit in the charging data were identified with this unsupervised learning approach; grouping stations in terms of their similarity.

IV. Implementation

A. Data Collection

EV charging opportunities are often grouped in three levels based on voltage, current and typical charging times. These levels are : Level 1 (L1), Level 2 (L2) and Level 3 (or DC Fast) [4]. Our study used real operational data from public electric vehicle charging stations provided by the New Brunswick Power Corporation. For this work, we selected the EV charging events that occurred between the dates of April 2019 and April 2020 at 37 Level-2 (L2) and 26 Level-3 (L3) public charging stations. The total number of charging events included in the analysis was 9,505. The total number of minutes spent recharging on L2 and L3 charging infrastructure was 551,635 minutes, which rounds up to 9,194 hours. The total amount of energy transferred to vehicles during the study period was 97,148.65 kWh.

Table I describes the features included in the raw EV charging data set. The raw data were fused with charging station location information and transformations were applied in order to feed the downstream processing.

TABLE I: Raw Data

Column Name	Description
Connection ID	Unique identifier for a connection
Recharge start time (local)	Timestamp denoting start of charging event
Recharge end time (local)	Timestamp denoting end of charging event
Account name	Unused (all null)
Card identifier	Unique identifier for a charging plan member
Recharge duration (hours:minutes)	Duration of charge event
Connector used	Connection used during charge event
Start state of charge (%)	State of charge percentage at beginning of charging event
End state of charge (%)	State of charge percentage after charging event is complete
End reason	Charge event end reason
Total amount	Unused (all null)
Currency	Unused (all null)
Total kWh	Energy transferred to vehicle during charging event
Station	Unique identifier for a charging station

B. Computational Environment

The software programs used in this work were packaged using Docker [7] containers in order to ensure a reproducible computational environment and to facilitate the distribution or extension of experimental workflows. A local Spark [8] cluster with 18 worker threads and 20GB of RAM was used to process the data. The data processing and analytical workflow was implemented using custom-written code, which used a standard scientific Python stack comprised of PySpark, Pandas, scikit-learn, NumPy and SciPy.

C. Analytical Workflow Implementation

Fig. 2 highlights noteworthy aspects of the analytical workflow implementation. We leveraged MLflow [9] to manage the development life-cycle of the proposed automated analytical workflow outlined in Fig. 1. The numbered boxes in Fig. 2 represent individual Spark jobs. The data flow is such that the output of one job is the input for the next job. Input and output file names contain parameter values that were used when calling the workflow's scripts. The grey elements represent a job's input file(s). The blue elements represent a job's output file(s).

Program elements were executed in sequence using Shell commands that called parameterized Python scripts. Examples of Bash scripts for executing workflow tasks 1 and 5 can be found here[1]. What follows is a description of each data processing workflow element from loading the initial raw data to applying the clustering algorithm. The detailed implementation of each task in the workflow are described as follows:

- **Task (1)**: We developed the ***one_way_hash.py*** script to import raw event data and cast column elements to appropriate types. Additionally, a one-way hash function was applied to the *Card identifier* column and the output was saved to a parquet file format.

[1]https://bitbucket.org/rr_mstrs/nb_ev_paper_1/src/master/

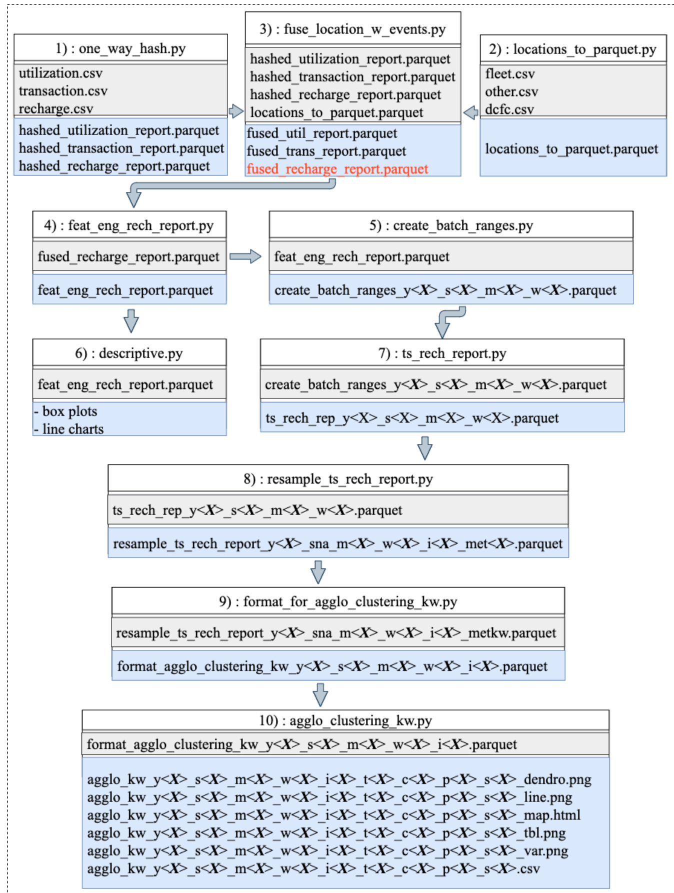


Fig. 2: Data Analytical Workflow Implementation.

- **Task (2)**: Then, the ***locations_to_parquet.py*** script is executed to import raw station location data and integrate multiple input files into one. The output is saved to a parquet file format.
- **Task (3)**: Next, the ***fuse_location_w_events.py*** script is triggered to fuse event data with the charging station location information.
- **Task (4)**: Focuses on recharge report event data in the downstream analysis. The ***feat_eng_rech_report.py*** script is used to create new features (contextualized) based on calculations involving existing data attributes and remove events with a duration of 5 minutes or less (eliminating 11% of the raw records).
- **Task (5)**: Extracts various partitions of the data using the ***create_batch_ranges.py*** script to enable analysis according to a particular week, month or season of the year.
- **Task (6)**: Produces descriptive analytics artefacts such as box plots and line charts using the ***descriptive.py*** script.
- **Task (7)**: Transforms event data from a row-based charging event summary format to a row-based, per-minute energy usage format using the ***ts_rech_report.py*** script.
- **Task (8)**: Re-samples event data with different aggregation periods using the ***resample_ts_rech_report.py*** script. Half-hour (for weekly clustering jobs), one-hour (for

monthly clustering jobs) and four-hour (for seasonal clustering jobs).

- **Task (9)**: Create a column-based, per-station and period energy usage format by pivoting the tables created in step (8) by executing the ***format_for_agglo_clustering_kw.py*** script.
- **Task (10)**: Executes the ***agglo_clustering_kw.py*** script to run the clustering algorithm using different input data partitions (e.g. weekly, monthly, seasonal).

Using different feature spaces, we can effortlessly run multiple experiments in our analytical workflow. Three experiments have been selected to evaluate our proposed analytical platform and their results are discussed in the next section (see Section V-B).

V. Results and Discussion

A. Descriptive Analytical Results

A descriptive analysis was performed in order to get a global understanding of the data. Fig. 3 summarizes the total monthly energy transfer to vehicles for both types of charging stations. The highest kWh month was August 2019 for both station types (L2 stations : 3,147.21 kWh, L3 stations 16,923.89 kWh). From Fig. 4 we observe, as would be expected, the aggregated monthly number of minutes spent charging at L2 stations is consistently higher than the same metric observed for L3 stations. From Fig. 5, we can observe the month with the highest number of L2 charging events was the month of January 2020. The peak month for L3 charging event frequency was August 2019, The high-frequency months for both station types were July-August and December-February, which largely corresponds to the summer and winter holiday seasons respectively.

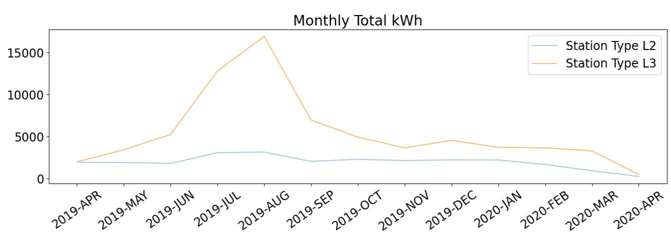


Fig. 3: Monthly kWh APR-2019 to APR-2020

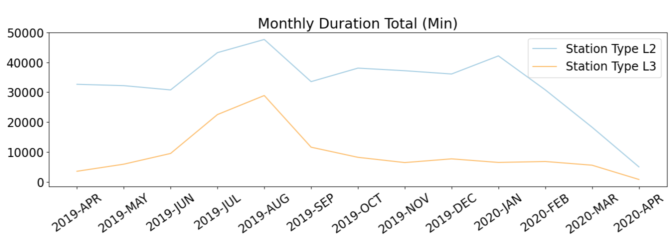


Fig. 4: Monthly Duration Minutes APR-2019 to APR-2020

In Figures 4 and 5 we observe a sharp decline in total monthly charge duration and event counts that becomes noticeable in February and continues into March. In March, the

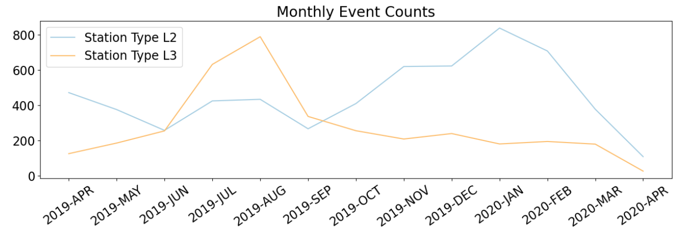
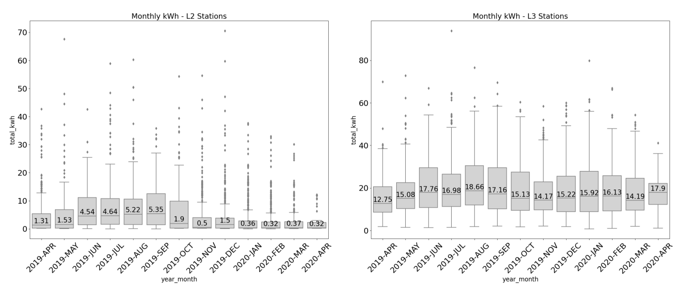


Fig. 5: Monthly Charge Events APR-2019 to APR-2020

provincial government made the first of many COVID-19-associated announcements related to school and other closures.

The box plots in Figs. 6a and 6b summarize the monthly kWh for charging events that occurred during the study time frame. The interquartile range (IQR) is calculated as the difference between quartiles 3 and 1. A common definition for an outlier is a value that is more than 1.5 times the interquartile range below the first quartile or above the third quartile. According to this definition, there were many outliers in the kWh values. This was especially true for L2 charge events. The points above or below the whiskers are values which are considered to be outliers in Figs. 6a and 6b.



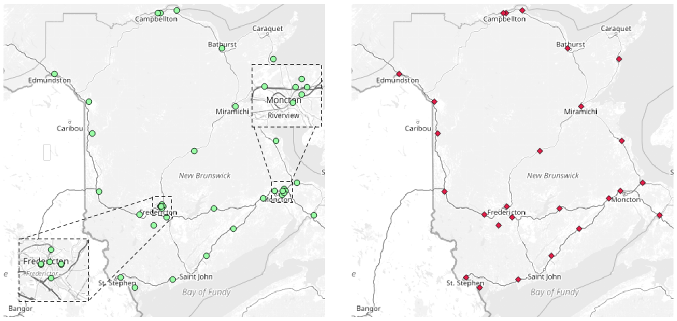
(a) L2 Stations (b) L3 Stations

Fig. 6: Monthly L2, L3 kWh APR-2019 to APR-2020

For both station types, most months had multiple outliers except for the months of June and September. L3 charging event kWh values have relatively less outliers when compared to L2 events. The charging events described in Figs. 3 through 6b were generated at public charging stations in New Brunswick, which are mapped in Fig. 7.

Public charging stations were strategically located throughout the province in order to allow for comfortable EV travel distances on the road network with ample access to charging opportunities. DC fast chargers (or L3 station outlets) generally transfer more energy to vehicles in a shorter amount of time when compared to L2 charging. This is apparent when comparing Tables II and III. The median monthly kWh values are consistently and significantly lower for L2 charging events.

There is significant variability in L2 charge event kWh values. The spread in these values is apparent when comparing the monthly mean and median values and when observing max and min values (See Table II). Additionally, the standard deviation is above 4.5 most months, which indicates that generally, on average, kWh values are above 4.5 units away



(a) L2 Stations (b) L3 Stations

Fig. 7: Charging stations in New Brunswick, Canada.

from the mean. A similar pattern of variability can be observed in L3 charging events (See Table III).

TABLE II: L2 Energy Transfer Basic Statistics - (kWh)

Y-M	N	Mean	Med	Std Dev	Min	Max
2019-APR	471	4.12	1.31	6.48	0.15	42.71
2019-MAY	375	5.08	1.53	7.59	0.05	67.55
2019-JUN	257	6.94	4.54	7.09	0.09	42.56
2019-JUL	424	7.26	4.64	8.27	0.0	58.87
2019-AUG	433	7.27	5.22	7.87	0.09	60.26
2019-SEP	266	7.66	5.35	7.33	0.0	35.79
2019-OCT	409	5.59	1.9	7.45	0.12	54.26
2019-NOV	619	3.45	0.5	6.15	0.13	54.62
2019-DEC	622	3.57	1.5	6.63	0.07	70.55
2020-JAN	837	2.63	0.36	4.69	0.01	37.62
2020-FEB	707	2.36	0.32	4.36	0.01	32.92
2020-MAR	377	2.45	0.37	4.4	0.03	30.13
2020-APR	108	2.22	0.32	3.38	0.16	12.18

TABLE III: L3 Energy Transfer Basic Statistics - (kWh)

Y-M	N	Mean	Med	Std Dev	Min	Max
2019-APR	125	15.93	12.75	10.89	1.76	69.86
2019-MAY	185	18.48	15.08	12.26	1.44	72.87
2019-JUN	254	20.64	17.76	12.75	1.32	66.92
2019-JUL	631	20.3	16.98	12.38	1.44	93.9
2019-AUG	788	21.48	18.66	12.42	1.82	76.61
2019-SEP	336	20.73	17.16	13.47	2.06	69.51
2019-OCT	255	19.35	15.13	13.3	1.65	60.36
2019-NOV	208	17.53	14.17	11.36	2.0	58.48
2019-DEC	239	19.01	15.22	13.35	1.78	59.99
2020-JAN	180	20.64	15.92	15.32	0.76	79.82
2020-FEB	194	18.8	16.13	13.15	0.96	66.88
2020-MAR	179	18.33	14.19	11.92	1.89	54.33
2020-APR	26	17.61	17.9	9.51	1.04	41.13

B. Diagnostics Analytical Results

This section highlights interesting outcomes from the cluster analysis of recharge events occurring at L2 and L3 charging stations across the province of New Brunswick for the study period. As discussed in Section IV-C, charging event data was partitioned using different time granularity and aggregation schemes. The results highlighted in this section used agglomerative clustering, PCA filtering at 70% variance with a 30 minute kWh aggregation rate. Additionally, longitude/latitude information was included as features for each station. Optimal dendrogram cut-offs were determined using the Caliński-Harabasz [10] method. Fig. 8 plots the weekly number of clusters observed for weekly time slices. From the figure, we observe the number of clusters varied significantly over the weekly periods and L3 charging stations generally had more clusters than L2 stations.

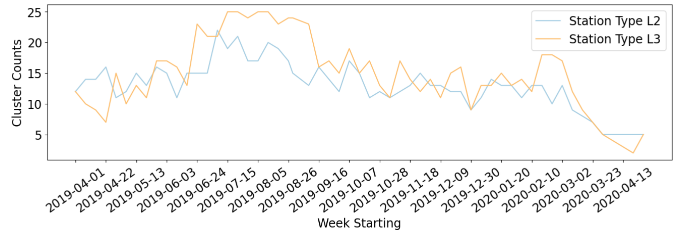


Fig. 8: Weekly Cluster Counts - L2 and L3 Events

1) Level 2 Recharge Events

First, we focus on L2 charging events and comment on clustering results for recharge events that occurred during the week starting on May 06, 2019 and compare these results to events that occurred in the week starting on August 05, 2019. During the week of May 06$^{\text{th}}$, the number of L2 recharge events was 68. Comparatively, the number of L2 recharge events in the week of August 05$^{\text{th}}$ was 104.

The dendrogram in Fig. 9 is a tree structure which contains all possible clusterings of a data set. The optimal cut-off determined using the Caliński-Harabasz [10] method for the week of May 06$^{\text{th}}$ was 2. This cut-off clustered stations in 12 groupings.

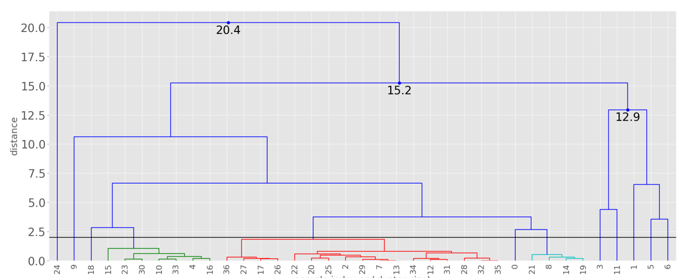


Fig. 9: Dendrogram - L2 Events - Week of May 06, 2019

The map in Fig. 10 provides a spatial view of the cluster members for the selected week. We see that the largest cluster (cluster 3 in green) is comprised of stations which are generally located in the center of the province. Additionally, cluster number 1 and 4 member stations seem to be located near the edges of the province only. Fig. 11 plots average kWh values for the top 3 largest clusters. As can be observed in this figure, cluster 4 stations, which generated the highest average kWh peaks, are located close to the province of Quebec and the State of Main in the United States. Cluster 1 stations are at the edge of Nova Scotia and Prince Edward Island. The largest cluster (cluster 3) represents consistently lower average kWh energy transfer patterns. Clusters 1 and 4 had relatively higher averages. Of note however, is the overall low aggregated kWh values for top cluster members for this week. The aggregated kWh of all L2 stations for the week was 334.62 kWh. The percentage of this total attributed to the top 3 largest cluster member stations was 12%. The smaller clusters consisted of member stations with higher aggregated kWh values.

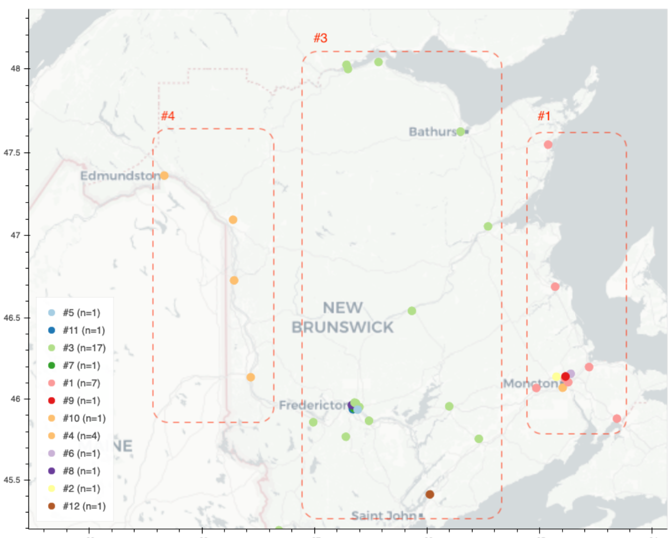


Fig. 10: Cluster Map - L2 Events - Week of May 06, 2019

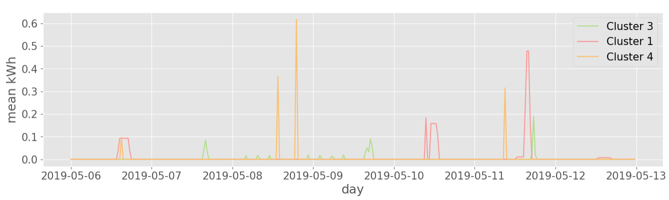


Fig. 11: L2 Events - Key Clusters - Week of May 06, 2019

We now focus on agglomerative clustering results for the 104, L2, recharge events that occurred during the week of August 05th. The number of events for this period was significantly higher when compared to the count of events during the week starting May the 06th. There were 65% more events in the week of August 05th. This increase in recharging events is reflective of the peak summer period which started in June for both station types and persisted throughout August (See Fig. 5). The top 3 largest clusters mapped in Fig. 12, generally included more member stations and largely covered more of the province to split it in two halves. There was one sub-cluster, cluster number 10, which was spatially included in cluster number 4, that included two stations situated near each other. The optimal cut-off in the dendrogram for this week was also 2, which resulted in 17 station groupings. The dendrogram for this clustering experiment is not included here for brevity.

Fig. 13 plots average kWh values for the top 3 largest station clusters observed on the week of August 05 th. As can be observed in this figure and in Fig. 12, cluster 10 member stations, which generated the highest average kWh peaks, contains two member stations located close to each other. Cluster 4 and 5 stations members generated similar average kWh peaks and roughly split the province in half.

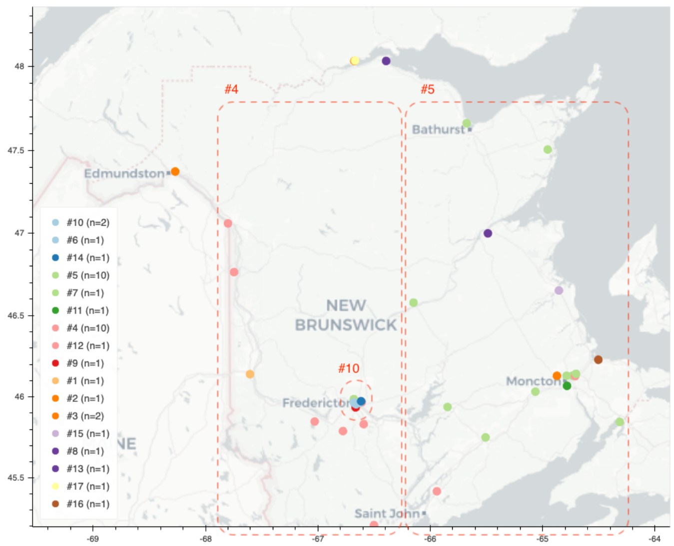


Fig. 12: Cluster Map - L2 Events - Week of August 05, 2019

The aggregated kWh values of top cluster members for this week were a little higher than values for the week of May 06th. Aggregated kWh of all L2 stations for this week was 879.53 kWh. The percentage of this total attributed to the top 3 largest cluster member stations was 17%. Again, the smaller clusters consisted of member stations with higher aggregated kWh values for the week.

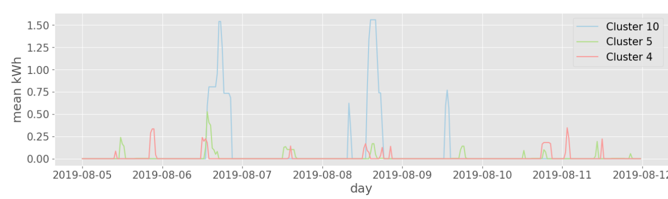


Fig. 13: L2 Events - Key Clusters - Week of August 05, 2019

2) *Level 3 Recharge Events*

As mentioned in the data transformation section (See IV-C), the analysis workflow can support running agglomerative clustering of weekly, monthly or seasonal batches. In this section, we focus on L3 stations and comment on station clusters generated from all recharge events that occurred during the month of May 2019. There were 185 recharge events that occurred at L3 stations during this month. The total amount of energy transferred to vehicles during the period was 3419.22 kWh. The agglomerative clustering experiment partitioned the L3 stations into 20 groupings. The largest cluster included 7 stations. All other clusters were comprised of single stations. The largest cluster (cluster number 9) was characterized as grouping stations which produced relatively low average kWh peaks. The aggregated value of kWh for all member stations in this cluster was 2% of the total for the month. As can be seen from Fig. 14, cluster number 9 member stations are broadly located in the top right quadrant of the province.

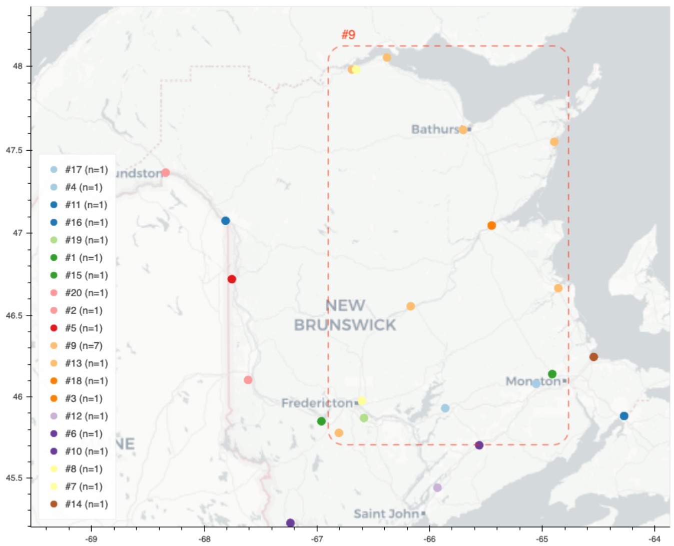


Fig. 14: Cluster Map - L3 Events - May, 2019

Fig. 15 plots average kWh values for the top 3 largest clusters. From this figure, we can observe that the cluster 9 member stations, on average, transferred very low amounts of energy to vehicles during the month.

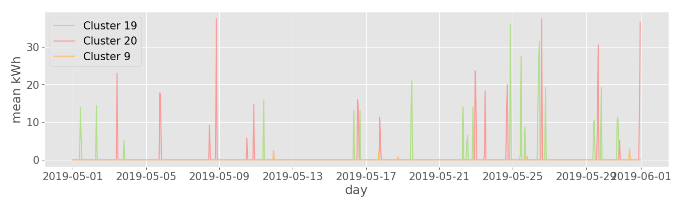


Fig. 15: L3 Events - May, 2019

Exploratory data analysis such as the one provided in Section V-A is a good start in acquiring an initial understanding of trends and visible patterns in a data set. However, the cluster analysis presented in this section can provide additional insights. Identifying under-utilized charging stations and similar stations groupings can be very useful to operators.

C. Discussion

The high capital costs of setting up public charging infrastructure and the use of public funds to support vehicle electrification necessitates robust informed decision making. The analysis in this work revealed that for some periods, sections of the province can be spatially divided into broad groupings of stations according to their energy utilization. The results highlighted in Section V demonstrate that agglomerative clustering is effective at grouping low kWh recharge stations together by considering spatial and temporal attributes. However, not all clustering experiments generated immediately observable and interesting results. Additionally, a manual inspection of clustering results revealed that stations with normal or relatively higher usage patterns were often not included in the same clusters. Additional clustering experiments with other algorithms and the filtering of outliers and fleet stations may provide additional insights by producing more compact and well-separated clusters.

VI. Conclusion and Future Research Work

A broad EV adoption scenario will require adequate public charging infrastructure. An understanding of EV charging patterns at public charging stations is crucial to foster adoption while managing costs and optimizing placement of charging infrastructure. The contributions in this work include an automated analytical workflow that enables the analysis of energy utilization patterns of public charging infrastructure using real charging data from station operators in New Brunswick. The outcomes of this research is believed to provide useful insights in planning and expanding infrastructure allocation. Future work will explore if state of charge can be used to make charge duration prediction for L3 chargers and whether this will be useful as a service to station operators and users. Additionally, future work will include exploring charging patterns from the point of view of user behavior.

VII. Acknowledgments

The authors would like to thank the New Brunswick Power Corporation for providing access to the EV charging data used in this research. This research was partially supported by the NSERC/Cisco Industrial Research Chair, Grant IRCPJ 488403-1.

References

[1] EVAG, "An electric vehicle roadmap for new brunswick a discussion document for public and stakeholder engagement," 2016.

[2] P. Ashkrof, G. Homem de Almeida Correia, and B. van Arem, "Analysis of the effect of charging needs on battery electric vehicle drivers' route choice behaviour: A case study in the Netherlands," *Transportation Research Part D: Transport and Environment*, vol. 78, p. 102206, Jan. 2020.

[3] E. Abotalebi, D. M. Scott, and M. R. Ferguson, "Why is electric vehicle uptake low in atlantic canada? a comparison to leading adoption provinces," *Journal of Transport Geography*, vol. 74, pp. 289–298, 2019.

[4] S. L. Monaca and L. Ryan, "The State of Play in Electric Vehicle Charging Services: Global Trends with Insight for Ireland," no. November, 2018.

[5] S. Hardman, A. Jenn, G. Tal, J. Axsen, G. Beard, N. Daina, E. Figenbaum, N. Jakobsson, P. Jochem, N. Kinnear, P. Plötz, J. Pontes, N. Refa, F. Sprei, T. Turrentine, and B. Witkamp, "A review of consumer preferences of and interactions with electric vehicle charging infrastructure," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 508–523, July 2018.

[6] H. Cao and M. Wachowicz, "An edge-fog-cloud architecture of streaming analytics for internet of things applications," *Sensors*, vol. 19, no. 16, p. 3594, 2019.

[7] C. Boettiger, "An introduction to docker for reproducible research," *ACM SIGOPS Operating Systems Review*, vol. 49, no. 1, pp. 71–79, 2015.

[8] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *HotCloud*, pp. 10–10, 2010.

[9] M. Zaharia, A. Chen, A. Davidson, A. Ghodsi, S. A. Hong, A. Konwinski, S. Murching, T. Nykodym, P. Ogilvie, M. Parkhe, *et al.*, "Accelerating the machine learning lifecycle with mlflow.," *IEEE Data Eng. Bull.*, vol. 41, no. 4, pp. 39–45, 2018.

[10] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.