



OPEN

DATA DESCRIPTOR

A City-scale and Harmonized Dataset for Global Electric Vehicle Charging Demand Analysis

Zihan Guo^{1,2,3}, Linlin You^{1,2}✉, Rui Zhu⁴, Yan Zhang⁵ & Chau Yuen⁶

With increasing policy and market support for electric vehicles (EVs) worldwide, analyzing EV charging demand is crucial for jointly optimizing transportation and energy systems. However, existing public datasets typically suffer from limited global coverage, coarse temporal resolution, and narrow feature availability. Here, we present CHARGED, a city-scale and harmonized dataset for global electric vehicle charging demand analysis. CHARGED contains hourly records from April 1 to September 30, 2023, covering about 12,000 charging chargers across six representative cities on six continents, including Amsterdam, Johannesburg, Los Angeles, Melbourne, São Paulo, and Shenzhen. Each entry encompasses core charging metrics (duration, volume, electricity price, and service price) alongside rich auxiliary information (weather variables, geospatial attributes, and multi-level static descriptors). CHARGED fills existing gaps and provides standardized data with spatiotemporal features aligned and multi-source information harmonized. Technical validation shows the potential of CHARGED to support in-depth characterization of user charging demand, and to impel the study of more advanced machine learning models, especially those enabling transfer learning across diverse urban contexts.

Background & Summary

With growing pressures on both urban air quality and energy security, electric vehicles (EVs) have become an important means of implementing low-carbon and sustainable mobility^{1,2}. Despite the accelerating global adoption of EVs, the development of charging infrastructure is not yet ready to empower cities in an optimal state that can fulfill the increasing EV charging demand cost-efficiently³. Hence, considerable strains on urban transportation and energy systems are emerging, as, on one hand, consumer expectations for charging convenience, speed, and reliability are continuing to rise, and on the other hand, grid operators and policymakers are struggling to ensure the balance between demand and supply⁴. Consequently, to forge a general research foundation, there is an urgent need for a scientific dataset that can systematically characterize EV charging behaviors from a global perspective. Such data are vital for advancing intelligence and autonomy in applications such as charging site location⁵, dynamic energy allocation⁶, and real-time traffic guidance⁷, as well as for guiding the development of urban infrastructure⁸, the upgrade of smart grids⁹, the evolution of intelligent transportation systems¹⁰, etc.

To enable these applications in reality, it becomes a key challenge about how to use EV charging data to make accurate forecasts. In general, traditional time-series methods, such as autoregressive integrated moving average (ARIMA)^{11,12}, offer certain advantages in supporting forecasting with highly regular patterns, and, accordingly, may become inefficient when confronted with nonlinear samples. To tackle these shortcomings, deep learning related approaches, especially with the fusion of spatial contexts and auxiliary features^{13–16}, have been widely studied, demonstrating superior capabilities in not only capturing localized charging behaviors but also modeling global dependencies across multiple temporal scales. In recent days, the foundation models adapted for time-series forecasting, through techniques such as prompt engineering¹⁷ or tokenization-based optimization¹⁸, have demonstrated outstanding out-of-the-box generalization capabilities to capture both fine-grained fluctuations and long-term trends by applying minimal architectural modification or parameter fine-tuning.

¹School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China. ²Guangdong Provincial Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China. ³Shanghai Innovation Institute, Shanghai, China. ⁴Institute of High-Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore, Republic of Singapore. ⁵Department of Informatics, University of Oslo, Oslo, Norway. ⁶School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, Republic of Singapore. ✉e-mail: lyou@ieee.org

Dataset	Spatial Coverage	Spatial Granularity	Time Spanning	Temporal Granularity	Record Count	Aux.
ACN-Data ³⁰	Three locations citywide	Site	N/A	N/A	30,000+	×
Residential EV ³¹	A location within a city	Charger	2018.12.01 – 2020.01.31	N/A	6,878	✓
EV Charging Fleet ³²	Several locations citywide	Charger	2018.05.11 – 2020.07.01	10 min	1,853,280+	×
Charging Transactions ³³	Several locations nationwide	Charger	2021.09.30 – 2022.09.30	15 min	72,856	✓
UrbanEV ³⁴	Urban regions of a city	Charger/Site	2022.09.01 – 2023.02.28	5 min/1 hour	1,194,600	✓
CHARGED (ours)	Urban regions of six cities	Charger/Site	2023.04.01 – 2023.09.30	5 min/1 hour	510,877,797	✓

Table 1. Comparison of representative public EV charging datasets. Note that N/A stands for information not explicitly provided, and Aux. represents whether there are any auxiliary features.

To support the study of these data-driven models, high-quality datasets become essential¹⁹. Although several public EV charging datasets have been released (with representative examples listed in Table 1), they can, in general, only support studies of models focusing on fusing information from a single or homogeneous source, and lack the ability to either test the adaptivity of related models in processing data with heterogeneous contexts, such as different cities, or enable transfer learning among these models to further improve their generalizability. Therefore, there is a need for a standard dataset that holds data from different cities with a high level of heterogeneity in economic development, mobility patterns, EV penetration, infrastructure completeness, and weather diversity²⁰.

Hence, we introduce CHARGED, a city-scale and harmonized dataset for global EV charging demand analysis. In general, CHARGED is prepared with three core objectives. First, it aims to fill existing data gaps by integrating EV charging data with auxiliary information from representative cities worldwide. Second, it intends to provide standardized data with spatiotemporal features aligned and multi-source information harmonized. Finally, it is prepared to provide a general foundation that provides necessary but rich content to support the study of how novel technologies, e.g., federated meta-learning²¹, and retrieval-augmented generation²², can be applied to advance the analysis of EV charging demand worldwide.

Methods

Data Overview. In general, CHARGED comprises 510,877,797 raw charging records collected from 11,953 charging chargers within a time range from April 1 to September 30, 2023. After data cleansing and harmonization, these records are aggregated to form hourly records for 4,280 sites generated through clustering of charging chargers. Besides EV charging data, i.e., duration, volume, electricity price (charges for electric vehicle charging), and service price (additional charges such as parking fees and idle fees), CHARGED also includes information about weather variables (temperature, precipitation, visibility, and other influencing factors), functional attributes (points of interest (POI) around a charging site, inter-site distances), and static descriptors (hierarchical data at charger, site, and city levels). By offering a unified, multi-scale, and richly annotated global dataset, CHARGED can support fine-grained analysis of EV charging behavior and enable the study of model generalizability that can impel the development of next-generation intelligent EV and energy management systems.

To facilitate a comprehensive analysis of global EV charging behaviors, six representative cities, one from each continent, have been selected based on multidimensional criteria including data availability (cities with continuous, richly attributed, and fine-grained charging records), geographic diversity (cities from each continent to ensure broad applicability), economic diversity (cities spanning various stages of economic development), and regional representativeness (cities with large populations, extensive charging site networks, and high electric vehicle adoption rates). The statistics for each city are outlined in Table 2. Raw data covers public EV charging chargers across the entirety of each city's urban area, collected from open online platforms for different regions over the half-year period from April 1 to September 30, 2023, including ChargePoint (<https://www.chargepoint.com/>), Chargefox (<https://www.chargefox.com/>), ChargePocket (<https://www.chargestations.co.za/cp/Index.aspx>), Chongdianba App (<https://apps.apple.com/cn/app/id1071506659>), and Tupi (<https://tupinambaenergia.com.br/>). These open platforms allow users to query the status of public EV charging chargers within the respective cities, retrieving attributes such as charging power, status, and electricity pricing, as well as geographic information including address, latitude, and longitude. Additionally, to further support the spatiotemporal analysis of urban EV charging behaviors, we also collected auxiliary data, e.g., POI within the boundaries of each city from OpenStreetMap (<https://www.openstreetmap.org/>), which includes details such as POI types and their corresponding geographic coordinates; and weather data from Visual Crossing API (<https://www.visualcrossing.com/>) containing meteorological site measurements for each city, which encompasses parameters such as temperature, precipitation, and visibility.

Data Cleansing. As shown in Fig. 1, to make the collected heterogeneous charging data ready for further usage, we design and implement the following data cleansing workflow:

1. **Parsing.** Raw charging records were parsed and converted into a standardized schema with predefined fields including timestamp, site identifier, instantaneous power, rated power, status code, and geographic coordinate. Charging duration and volume were generated according to the plug-level status and power. As records for each charging charger may host multiple charging plugs, the total charging duration at each timestamp was calculated as the number of plugs labeled as “charging”. The instantaneous power of each plug at each timestamp was collected and multiplied by its corresponding duration to calculate the charging volume. However, since the raw data cannot fully capture transient power fluctuations, this calculation

City	Continent	Country	Abbreviation	Land Area	Population	Record Count	Charger Count	Site Count
Amsterdam	Europe	Netherlands	AMS	166	931,298	244,600,416	8880	2449
Johannesburg	Africa	South Africa	JHB	1,643	6,198,000	11,335,392	108	47
Los Angeles	North America	United States	LOA	2,629	8,092,394	53,851,104	521	229
Melbourne	Australia	Australia	MEL	4,264	3,848,745	14,154,624	80	63
São Paulo	South America	Brazil	SPO	1,521	11,451,999	6,235,204	55	47
Shenzhen	Asia	China	SZH	1,748	17,791,600	180,701,057	2309	1445

Table 2. Summary of six worldwide cities in CHARGED. For each city, it shows land area (km^2 , calculated based on the city boundary data), population, as well as the total count of collected records, charging chargers and charging sites. Note that charging sites are generated via geographical clustering of recorded charging chargers.

- may introduce some bias. Therefore, a fine-grained sampling frequency was implemented, with the rated power substituted in the absence of instantaneous power, in order to more closely approximate the actual charging volume. It should be noted that the use of rated power occurs only in a subset of records from MEL, representing a negligible proportion. Since our downstream processing pipeline includes outlier detection and correction, no additional adjustment factor was applied to these instances. Additionally, for records lacking information about electricity price or service price, they were filled according to the pricing rules retrieved from the raw data. For chargers without any service price, a value of zero was used.
- Filtering.** Data filtering was made both spatially and temporally. First, geographic coordinates were re-projected to a unified reference system, and any records located outside the administrative boundary of their respective city were excluded. Second, as raw records may include historical entries outside the target period, only records within the study window were retained. It should be noted that capturing seasonal characteristics typically requires a full year of data. Due to limitations in data completeness and quality, CHARGED covers only a six-month period, which may limit its ability to capture seasonal patterns, especially in regions with pronounced temperature variations. Nevertheless, this time window still effectively reveals variations in charging behavior and supports tasks such as short-term and long-term forecasting, as well as cross-domain knowledge transfer. The corresponding validations are presented in the following section.
 - Denoising.** To mitigate data quality issues, we implemented an automated denoising procedure that harmonizes both spatial inconsistencies and numerical anomalies. As for spatial inconsistencies, e.g., a charging charger may appear under slightly different latitude/longitude values at different timestamps, we calculated the mean of observed latitude and longitude for each charging charger, and then used this centroid as its geographic location. Meanwhile, as for numerical anomalies, the implausible measurements of charging duration or volume (e.g., negative values) were first set to missing. Then any remaining values lying more than four times the interquartile range below the first quartile or above the third quartile were identified as outliers²³ and likewise imputed as missing.
 - Imputation.** To address missing values flagged during denoising as well as those inadvertently omitted during data collection, we designed an imputation strategy operating at the individual charging charger level. Using timestamps as the index, the temporal interpolation is implemented by filling approximate values calculated according to the index and the actual value around this interpolation point. For long missing sequences or boundary segments, where direct interpolation is infeasible, we applied linear interpolation with subsequent forward- and backward-filling, thereby ensuring seamless continuity and completeness of the dataset. The two imputation methods account for 10.70% and 0.99% of the total data, respectively.

Data Harmonizing. As illustrated in Fig. 1, to achieve data unification across diverse cities on multiple continents, the cleansed data was further harmonized in three steps to ensure data consistency:

Temporal Alignment. Due to network latency, recorded sampling times exhibited slight deviations, so raw timestamps were rounded down to the nearest five-minute interval. Any duplicate records for the same charger and timestamp were merged by taking the mean of related metrics such as charging duration and volume. Although Coordinated Universal Time (UTC) is typically used for temporal alignment²⁴, we retained each city’s local time zone to better capture its intrinsic temporal distribution. To further smooth high-frequency noise, the five-minute-resolution data were then aggregated into hourly bins by summing parameters such as charging duration and volume and averaging variables like electricity price and service price.

Spatial Formalization. To better represent real-world EV charging infrastructure layout, we defined the virtual charging site as a spatial cluster of neighboring physical charging chargers. Accordingly, we implemented an adaptive spatial aggregation strategy to group points of charging chargers and generate virtual sites that consolidate geographic information into coherent units. To be specific, the primary task was to determine the rules for defining neighboring charging chargers within a charging site. Consequently, by calculating pairwise geodesic distances between all charging chargers, we created a one-dimensional distance array for each city. By sorting this array, its largest jump in first-order differences was used as the adjacent radius. This procedure produced a mean radius of 47.37 meters across the six cities, including 32.96 meters in MEL, 34.97 meters in SPO, 40.29

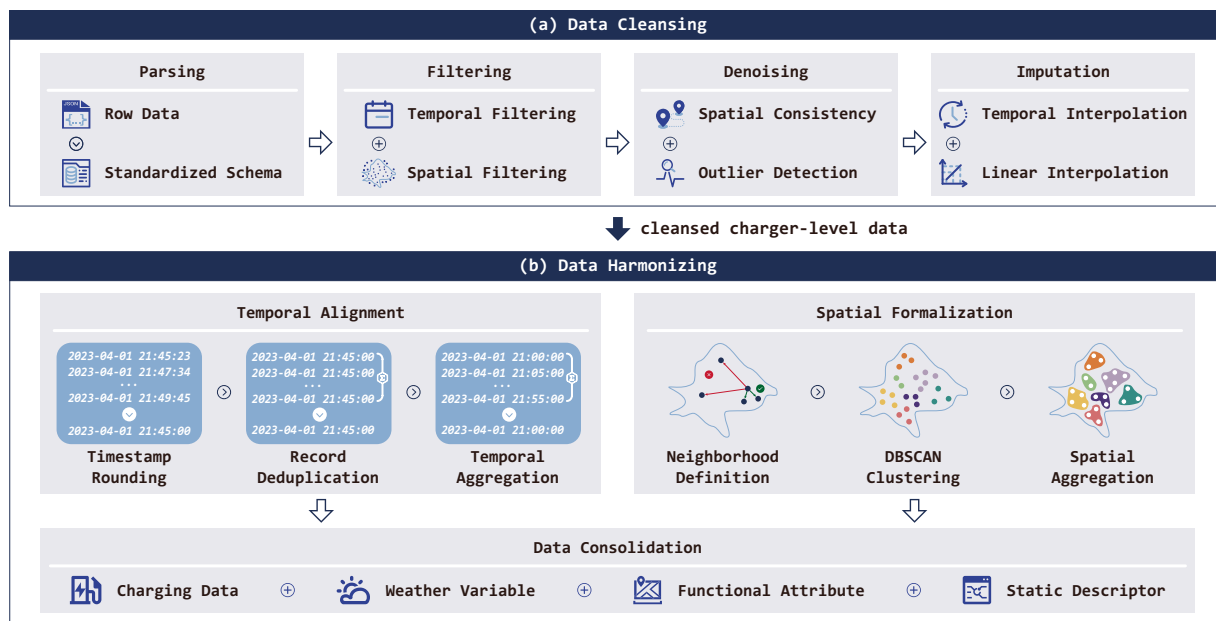


Fig. 1 Overall workflow for data cleansing (a) and data harmonizing (b). Raw records are cleansed by successively applying parsing, filtering, denoising, and imputation. The cleansed charger-level data are then harmonized into a unified site-level dataset through temporal alignment, spatial formalization, and data consolidation. This figure was created with graphic elements provided by Iconfont (<https://www.iconfont.cn>).

meters in JHB, 41.95 meters in SZH, 47.36 meters in AMS, and 86.70 meters in LOA. For each city, we then applied the DBSCAN algorithm²⁵ by using its corresponding radius as the parameter to determine connectivity to cluster charging chargers into charging sites. Note that the geolocation of each generated charging site is the mean of its cluster members' latitudes and longitudes, with geographic locations in the city as shown in Fig. 2. Furthermore, within each charging site, charging duration and volume were aggregated by summation, while electricity and service price were averaged to prepare the hourly record.

Data Consolidation. Data for each city were structured into four major modules, i.e., charging data, weather variables, functional attributes, and static descriptors. Charging data, including duration, volume, electricity price, and service price, were reorganized into separate matrices indexed by hourly timestamps and charging site identifiers after temporal and spatial harmonization. Then, city-level weather variables (e.g., temperature, precipitation, and visibility) were normalized on the same hourly grid to ensure perfect temporal alignment with the charging data. Meanwhile, functional attributes, represented by POI and their inter-site distance matrix, were prepared, which are time-invariant and shared across all sites within each city. Finally, static descriptors consisted of statistical indicators at the charger, site, and city levels, which were computed only for entities with valid data and stored in flat-table format. Through the cross-mapping and spatiotemporal alignment of these four components, we produced a coherent, semantically unified, and globally consistent dataset. Related statistics of some key features are provided in Table 3.

Data Records

The dataset is available at GitHub (<https://github.com/IntelligentSystemsLab/CHARGED>) and Zenodo²⁶. City-specific datasets are organized in separate directories, and each city directory contains two versions of the data, i.e., one complete and one excluding charging sites with zero duration and volume (given that it is inactivity due to no charging demand). All kinds of data are saved in Comma Separated Values (*.csv) format for ease of use. A description, including file overviews and data field definitions, is provided as a Markdown (*.md) file in the root directory. In addition, we supply a comprehensive suite of Python (*.py) scripts that provide standardized interfaces and illustrative examples for data preprocessing, model training, and deployment. In general, following files are included in CHARGED:

1. (duration.csv, volume.csv, e_price.csv, and s_price.csv) provide site-level hourly charging data, including charging duration (measured in hours), volume (measured in kilowatt-hours), electricity price, and service price. Prices are recorded in the local currency of each city without unit conversion.
2. (weather.csv) contains nineteen city-level meteorological variables at hourly resolution, all expressed in metric units.
3. (distance.csv) is part of functional attributes, and represents the distance matrix computed via a WGS-84 ellipsoidal geodesic algorithm, where both rows and columns are indexed by unique site identifiers, yielding a symmetric matrix in kilometers.
4. (poi.csv) is part of functional attributes to store the collected POI in each city. It contains POI types and

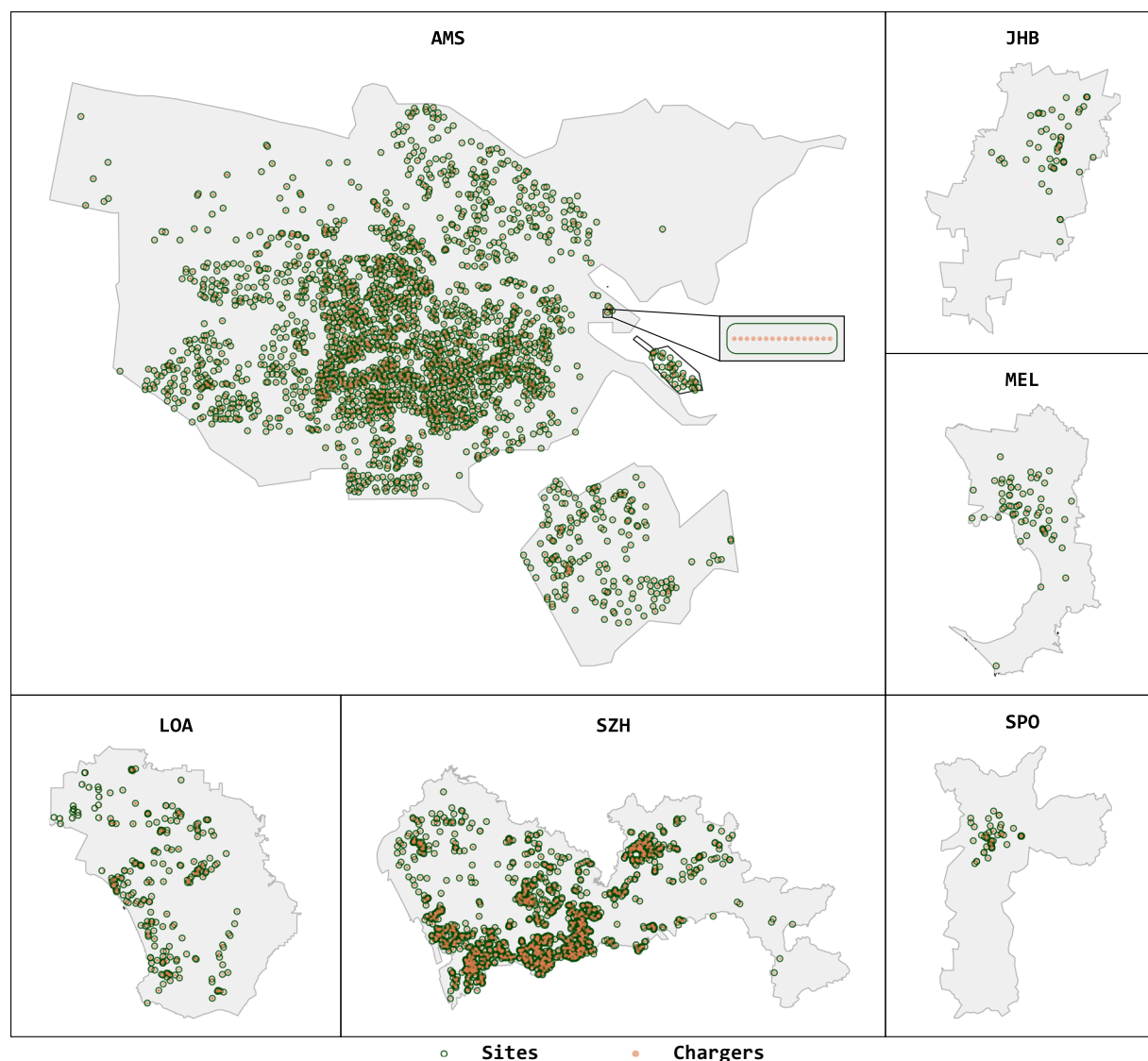


Fig. 2 Spatial distribution of charging chargers and the generated charging sites across six cities. Charging chargers are shown as orange filled circles, while charging sites are indicated by green hollow circles. To illustrate the clustering relationship in detail, an inset for AMS highlights one site composed of a row of 16 charging chargers, demonstrating the rationale behind clustering them as a single charging site. It should be noted that most cities have full-city coverage in CHARGED. However, for a few cities (such as SPO and JHB), CHARGED currently includes data only for some lower-level administrative regions. Consequently, we also employ city-level maps for visualization here and will supplement the dataset once the full data become available.

geographic coordinates. The types cover all subcategories under the major OpenStreetMap features, and the provided coordinates enable users to flexibly explore various POI integration strategies, such as spatial buffering, association matching, and filtering criteria.

- (chargers.csv, sites.csv, and info.csv) provide static descriptors at three hierarchical levels, i.e., charger, site, and city. It includes unique identifiers, hierarchical relationships, geospatial data, and charging statistics, all stored in flat-table format.
- (README.md) offers a comprehensive dataset overview located in the root directory, including file descriptions, field names, data types, units, and semantic definitions.
- (*py scripts) implement standardized interfaces and illustrative examples organized into *api/* and *example/* subdirectories in the root directory.

Technical Validation

To validate the efficiency and effectiveness of CHARGED for EV charging demand analysis, we conduct a multi-perspective evaluation using hourly charging volume forecasting as a representative task. Charging volume is selected as the evaluation metric because, compared to charging duration, it better captures power fluctuations and energy demand during the charging process, and more effectively supports downstream application

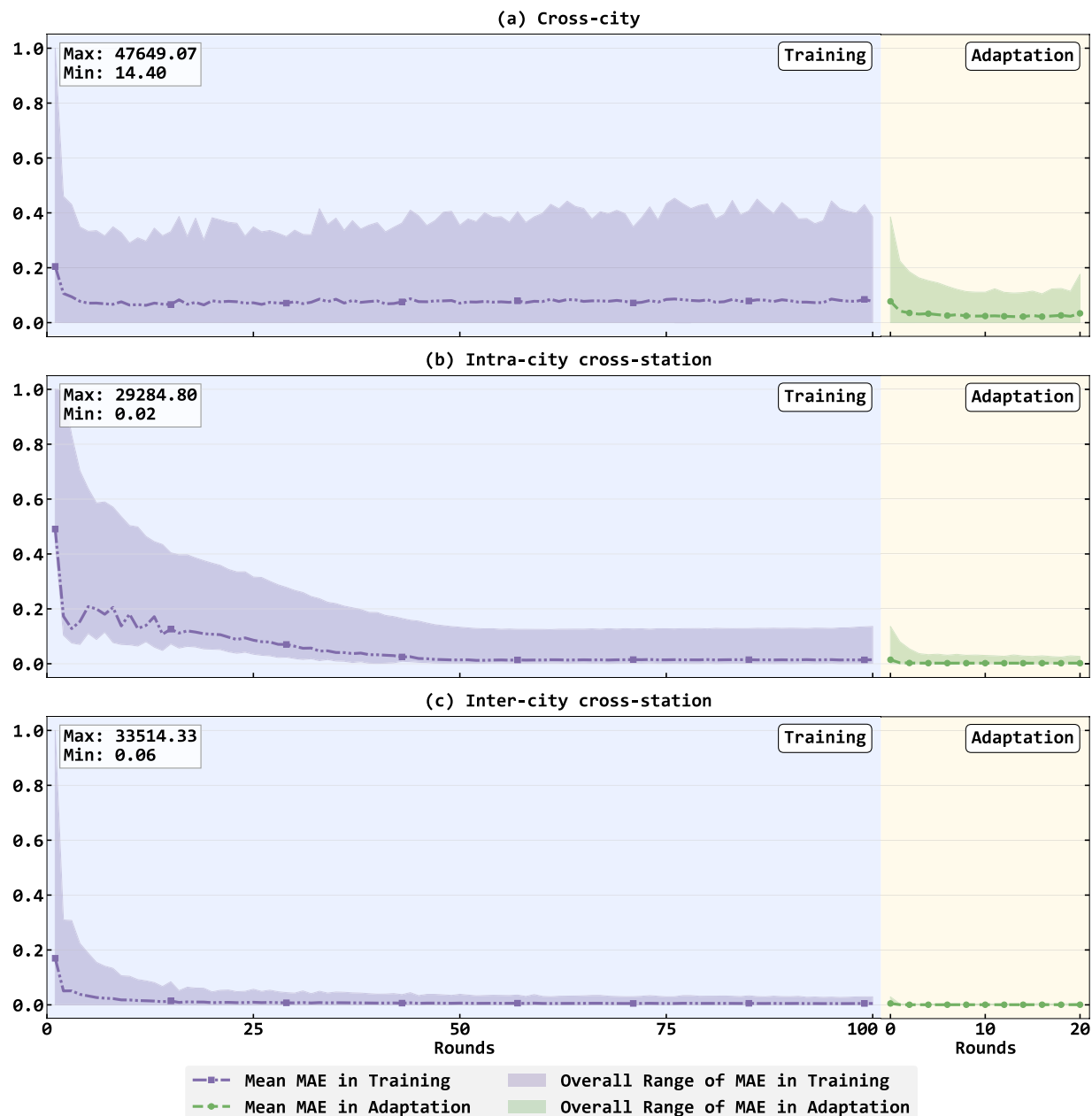


Fig. 3 Performance under three knowledge transfer scenarios using CHARGED, including cross-city transfer (a), intra-city cross-site transfer (b), and inter-city cross-site transfer (c). The changes in MAE were tracked during both training and adaptation phases. The curves were min-max normalized, with the corresponding extrema annotated in the top-left corner of each subplot. The results demonstrate pronounced reductions in MAE during both training and adaptation, thereby validating the efficacy of CHARGED in supporting transfer learning.

scenarios. The core objectives include the following four aspects, namely, 1) verification of its support for demand forecasting, 2) evaluation of performance gains imparted by its auxiliary features, 3) exploration of its capacity to support knowledge transfer, and 4) investigation of its usage on foundation models. Through these tasks, we aim to showcase the high quality and practical value of CHARGED.

In total, fifteen forecasting models were evaluated, encompassing the following three categories, namely 1) traditional statistical time-series models, including the autoregressive model (AR)¹¹ and ARIMA¹² methods; 2) novel deep-learning models, including FreTS¹³, ConvTimeNet (CTN)¹⁴, SegRNN¹⁵, and MultiPatchFormer (MPF)¹⁶, which are designed based on model architectures spanning from multilayer perceptrons (MLP), convolutional neural networks (CNN), recurrent neural networks (RNN), to Transformer; and 3) time-series forecasting foundation models drawn from three leading families consisting of Amazon's Chronos-T5²⁷, Salesforce's Moirai-1.1-R²⁸, and AutonLab's Moment-1²⁹, with three model sizes picked from each family.

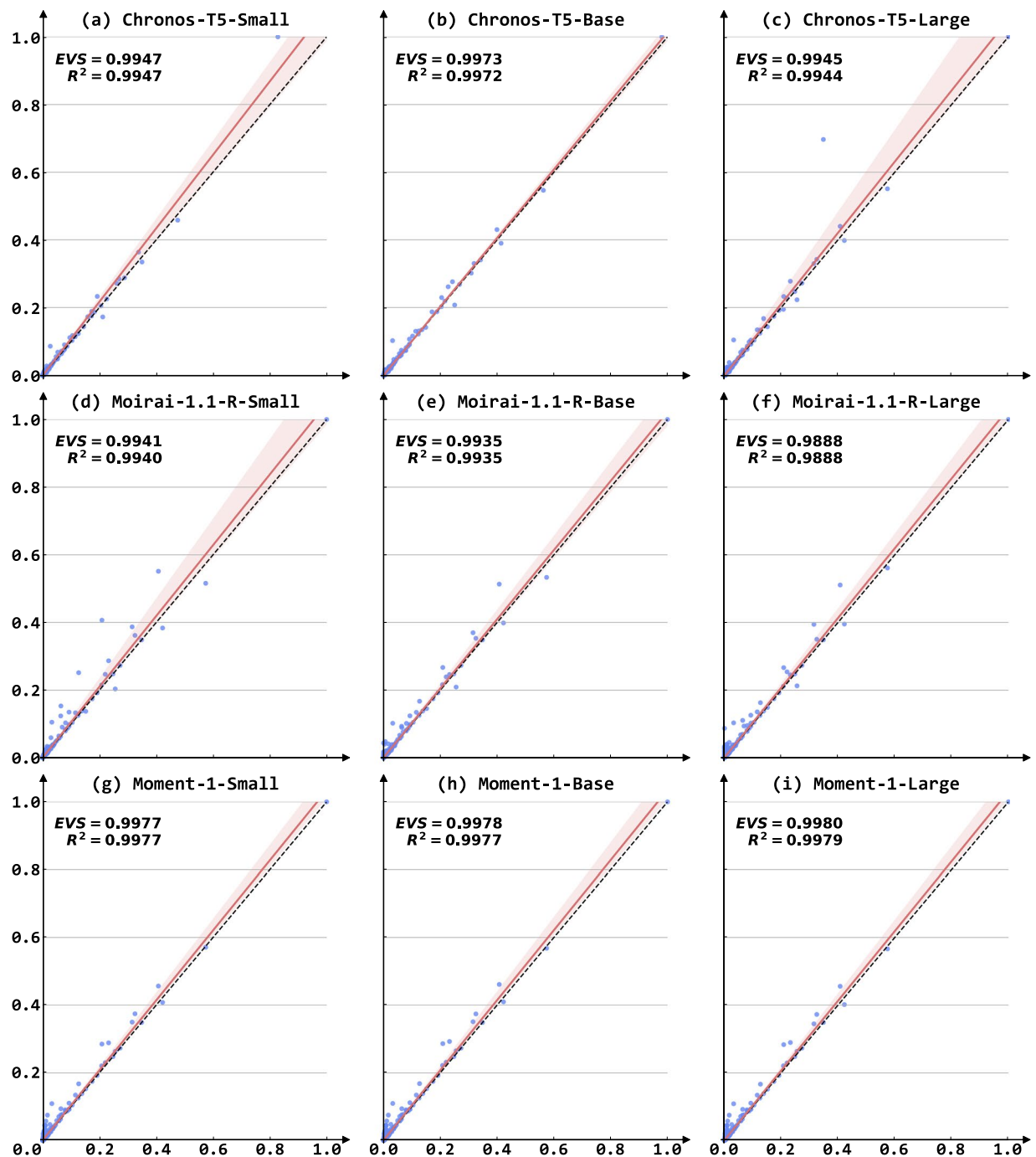


Fig. 4 Regressions of predicted versus actual values for long-term charging volume forecasting. In each subplot, the red solid line denotes the linear regression fit, the black dashed line marks the ideal $y = x$ reference, and the shaded band represents the regression confidence interval. EVS and R^2 are annotated to quantify predictive accuracy. Subplots (a–c), (d–f), and (g–i) correspond to the Chronos-T5, Moirai-1.1-R, and Moment-1 model families, respectively, each evaluated at small, base, and large scales for zero-shot one-month-ahead charging-volume prediction based on five months of historical data. The evident zero-shot capabilities of these foundation models further illustrate the rich temporal structure captured by CHARGED.

To ensure fairness, the evaluation was conducted under identical experimental settings and fixed random seeds. Moreover, model performance was measured using six widely accepted regression metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Median Absolute Error (MedAE), R-squared (R^2), and Explained Variance Score (EVS).

Results to Support Charging Demand Forecasting. This evaluation component was specifically designed to assess short-term forecasting performance. Data here were split into training, validation, and test

Level	Feature	Unit	Max	Min	Mean	Median	Q1	Q3	Std
Site	Duration	Hour	439.65	0.00	3.93	1.17	0.58	3.00	10.13
	Volume	Kilowatt-hour	93592.24	0.00	85.10	8.17	4.01	22.17	1427.39
	Perimeter	Meter	1002.50	230.33	364.20	398.91	275.33	399.31	108.38
	Area	Square Meter	67588.70	4206.33	10807.68	11576.35	6013.50	11593.27	7061.62
	Average Power	Kilowatt	360.00	0.00	10.35	6.60	6.40	7.00	23.04
City	Temperature	Degree Celsius	37.00	− 2.20	18.37	17.80	13.60	23.00	6.67
	Precipitation	Millimeter	107.51	0.00	0.08	0.00	0.00	0.00	1.36
	Visibility	Kilometer	45.50	0.00	14.16	11.40	10.00	16.00	7.86
	Site Distance	Kilometer	93.87	0.03	12.60	8.00	4.17	18.79	11.47
	POI Count	—	617806.00	42782.00	142719.33	48603.50	47257.75	50934.75	212481.62

Table 3. The statistical summary of multi-scale features. At the charging site level, charging duration, volume, average power, cluster-derived coverage perimeter, and coverage area are reported. At the city level, temperature, precipitation, visibility, distance among sites, and the number of POI are included. For each feature, the maximum, minimum, mean, median, first and third quartiles, and standard deviation are presented.

subsets in a ratio of 8:1:1. The models incorporate a 12-hour input window (12 time steps) to forecast the value in the subsequent hour. Moreover, they were trained for 50 rounds and evaluated via a month-based six-fold cross-validation strategy, with metrics reported as the average across all test folds.

Table 4 presents the forecasting performance of different models across multiple cities. Notably, traditional time-series models such as AR and ARIMA, which rely solely on historical charging volume, consistently underperformed. This highlights the inadequacy of linear autoregressive and statistical approaches in capturing the complex charging behavior observed in modern urban environments. In contrast, models like CTN and SegRNN, which incorporate spatial information and support nonlinear temporal modeling, demonstrate significant performance gains. These results confirm the effectiveness and robustness of using spatiotemporal information to support demand forecasting. Going further, advanced models, such as MPF, which leverages a multi-scale segmented temporal modeling strategy, and FreTS, which incorporates frequency-domain transformations, achieved the best or second-best results. Their ability to capture trends under scenarios with high volatility highlights their strength in modeling oscillatory patterns in time series. In summary, CHARGED's hourly spatiotemporal resolution provides a rich foundation for forecasting tasks. The dataset not only supports robust demand forecasting but also enables effective nonlinear spatiotemporal modeling. Future work may explore the development of dedicated modules that can better capture spatiotemporal dependencies hidden in charging behavior to further improve forecasting accuracy and generalization capability of related models.

Usage of Auxiliary Feature. The same training and evaluation setup here was adopted as the first experiment, except that only FreTS is used as the representative model to investigate the influence of auxiliary features on prediction performance. Moreover, seven auxiliary feature configurations were considered in this evaluation. First, the baseline setting (denoted as None) includes no auxiliary features. Second, one of the following features, i.e., electricity price (p_e), service price (p_s), temperature (T), precipitation (P), and visibility (V), is separately added to the baseline setting, which forms five feature-specific settings. Finally, a full setting (All) indicates all above five auxiliary features are used together with the baseline setting.

The results across different cities and configurations are summarized in Table 5. It demonstrates that the inclusion of auxiliary features generally leads to substantial improvements in forecasting performance across most cities, indicating their practical relevance in influencing EV charging behaviors. Notably, as the available weather data are currently limited to the city level, they may not accurately capture local meteorological conditions at individual charging sites. While some performance improvements can be observed, finer spatial-resolution weather data would undoubtedly enhance demand analysis, and it will be updated into CHARGED accordingly if such data become available. Furthermore, different cities exhibit varying sensitivities to specific features. For instance, the inclusion of P yields the best performance for AMS, while p_e leads to the best results across all metrics in LOA, reflecting the heterogeneity of charging behavior patterns among cities. Interestingly, in some cases, incorporating all auxiliary features results in performance degradation, highlighting the importance of proper feature selection and modeling despite the availability of this rich information. Overall, the extensive spatial, economic, and behavioral descriptors in CHARGED provide strong support for advanced nonlinear and multimodal spatiotemporal modeling.

Potential to Support Knowledge Transfer. In this evaluation, we employed federated learning, which enables collaborative and privacy-preserving learning among data owners, to assess the ability of CHARGED to support transfer learning. We assumed that each data source was a client and created three testing scenarios, i.e., 1) in the cross-city transfer scenario, each city served as a client, with SZH designated as the test client and the remaining cities as training clients; 2) in the intra-city cross-site transfer scenario, all charging sites within SZH were partitioned into training and test clients at an 8:2 ratio; and 3) in the inter-city cross-site transfer scenario, 30 sites were selected from each city and, then split into training and test clients at an 8:2 ratio. For all scenarios, models were trained for 100 rounds and then fine-tuned over 20 adaptation rounds per test client, using 50% of the test client's data for fine-tuning and the remaining 50% for evaluation. Note that reported metrics are the averaged performance across all test clients.

AMS						
Model	MAE	RMSE	RAE	MedAE	R ²	EVS
AR	7.492	25.520	0.197	6.089	0.781	0.783
ARIMA	7.573	25.650	0.199	6.197	0.780	0.781
CTN	3.720	10.410	<u>0.078</u>	<u>1.599</u>	<u>0.950</u>	<u>0.954</u>
SegRNN	3.230	10.690	0.114	2.098	0.929	0.931
MPF	1.461	<u>5.567</u>	0.052	0.814	0.978	0.978
FreTS	<u>2.757</u>	3.372	0.556	2.752	0.625	0.805

LOA						
Model	MAE	RMSE	RAE	MedAE	R ²	EVS
AR	4.834	8.304	0.699	4.617	0.618	0.638
ARIMA	5.278	9.019	0.764	5.134	0.545	0.573
CTN	1.767	3.441	0.254	1.249	0.934	0.935
SegRNN	3.001	5.454	0.433	2.621	0.835	0.837
MPF	<u>1.536</u>	<u>3.314</u>	<u>0.222</u>	<u>0.917</u>	<u>0.939</u>	<u>0.939</u>
FreTS	1.503	3.259	0.217	0.861	0.941	0.942

SPO						
Model	MAE	RMSE	RAE	MedAE	R ²	EVS
AR	3.168	6.668	0.687	2.929	0.211	0.228
ARIMA	3.386	7.386	0.733	3.145	0.056	0.068
CTN	1.308	3.642	0.281	0.881	0.761	<u>0.767</u>
SegRNN	1.206	3.676	0.274	0.753	0.746	0.747
MPF	<u>0.997</u>	<u>3.597</u>	<u>0.221</u>	<u>0.505</u>	<u>0.766</u>	0.767
FreTS	0.952	3.485	0.211	0.444	0.780	0.781

JHB						
Model	MAE	RMSE	RAE	MedAE	R ²	EVS
AR	4.861	9.834	0.714	4.140	0.122	0.147
ARIMA	5.044	10.184	0.740	4.356	0.110	0.130
CTN	1.537	4.094	0.223	1.024	0.825	0.829
SegRNN	1.396	4.195	0.191	0.935	0.819	0.823
MPF	<u>0.915</u>	<u>3.707</u>	<u>0.136</u>	<u>0.345</u>	<u>0.851</u>	<u>0.851</u>
FreTS	0.848	3.674	0.127	0.266	0.853	0.853

MEL						
Model	MAE	RMSE	RAE	MedAE	R ²	EVS
AR	30.009	106.953	0.555	27.227	0.440	0.443
ARIMA	31.388	113.155	0.579	27.492	0.355	0.364
CTN	16.138	71.231	0.299	11.940	0.750	0.751
SegRNN	24.806	89.869	0.458	21.718	0.575	0.576
MPF	<u>15.999</u>	73.255	<u>0.296</u>	<u>11.638</u>	0.735	0.735
FreTS	15.282	<u>71.575</u>	0.283	10.824	<u>0.745</u>	<u>0.745</u>

SZH						
Model	MAE	RMSE	RAE	MedAE	R ²	EVS
AR	499.497	7029.149	0.521	488.559	0.817	0.817
ARIMA	128.656	628.770	0.134	121.011	0.980	0.980
CTN	<u>50.495</u>	<u>261.486</u>	<u>0.054</u>	37.771	<u>0.997</u>	<u>0.997</u>
SegRNN	68.385	302.189	0.071	57.252	0.995	0.995
MPF	47.600	247.707	0.050	34.496	0.997	0.997
FreTS	51.266	278.575	0.055	<u>36.041</u>	0.996	0.996

Table 4. Performance of charging volume forecasting based on CHARGED. It includes results for two traditional statistical time-series models and four novel deep-learning models across six representative cities. Models, which can jointly leverage temporal and spatial features, consistently outperformed their counterparts, highlighting the pronounced spatiotemporal dependencies inherent in CHARGED.

Figure 3 illustrates performance curves during the training and adaptation phases across three knowledge-transfer scenarios. Overall, MAE decreases markedly with the increase of training rounds, and stabilizes rapidly after only a few adaptation rounds. Although the test clients are not the same, cross-scenario comparisons can generally reveal that more granular and hierarchical transfer strategies produce smoother convergence curves, indicating that global models can more effectively capture generalizable features. However, as adaptation rounds continue, some scenarios exhibit performance fluctuations and minor degradations, showing a sign of overfitting. This experiment demonstrates the value and potential of CHARGED produced for heterogeneous knowledge transfer, and reveals the possibility of integrating with novel technologies to further improve the forecasting capability of models as well as their adaptability in better handling heterogeneous scenarios.

AMS						
Auxiliary	MAE	RMSE	RAE	MedAE	R ²	EVS
None	2.757	<u>3.372</u>	0.556	2.752	0.625	0.805
p_e	<u>2.542</u>	3.389	0.546	<u>2.493</u>	0.625	<u>0.824</u>
p_s	2.542	3.389	0.546	2.493	0.625	0.824
T	2.609	3.465	<u>0.534</u>	2.591	0.625	0.814
P	2.166	2.943	0.489	2.122	0.677	0.833
V	2.727	3.592	0.549	2.650	0.624	0.812
All	2.757	3.612	0.556	2.732	<u>0.661</u>	0.810

LOA						
None	1.503	3.259	0.217	0.861	0.941	0.942
p_e	1.432	3.210	0.207	0.785	0.943	0.943
p_s	<u>1.432</u>	<u>3.210</u>	<u>0.207</u>	<u>0.785</u>	<u>0.943</u>	<u>0.943</u>
T	1.522	3.242	0.220	0.919	0.942	0.943
P	1.434	3.213	0.207	0.786	0.943	0.943
V	1.461	3.242	0.211	0.822	0.942	0.942
All	1.472	3.225	0.213	0.850	0.943	0.943

SZH						
None	51.266	278.575	0.055	36.041	0.996	0.996
p_e	<u>47.903</u>	<u>256.004</u>	<u>0.051</u>	34.275	<u>0.997</u>	<u>0.997</u>
p_s	47.903	256.004	0.051	34.275	0.997	0.997
T	50.366	276.360	0.054	35.194	0.996	0.996
P	49.029	271.646	0.052	<u>33.660</u>	0.996	0.996
V	49.005	269.788	0.052	34.021	0.996	0.996
All	44.600	225.731	0.048	32.978	0.997	0.997

JHB						
None	<u>0.848</u>	3.674	<u>0.127</u>	<u>0.266</u>	0.853	0.853
p_e	0.961	3.711	0.141	0.403	0.851	0.851
p_s	— [*]	—	—	—	—	—
T	0.902	<u>3.660</u>	0.134	0.365	<u>0.854</u>	<u>0.854</u>
P	0.866	3.666	0.128	0.301	0.853	0.854
V	0.832	3.667	0.123	0.258	0.853	0.853
All	0.923	3.648	0.137	0.374	0.856	0.856

MEL						
None	15.282	<u>71.575</u>	0.283	10.824	<u>0.745</u>	<u>0.745</u>
p_e	<u>15.127</u>	72.768	0.281	<u>10.549</u>	0.735	0.736
p_s	—	—	—	—	—	—
T	15.305	73.400	0.285	10.767	0.726	0.726
P	15.106	72.842	<u>0.281</u>	10.423	0.733	0.734
V	15.241	72.634	0.283	10.800	0.733	0.734
All	15.138	70.847	0.280	10.649	0.752	0.752

SPO						
None	<u>0.952</u>	3.485	<u>0.211</u>	0.444	0.780	0.781
p_e	0.970	3.469	0.215	0.479	0.782	0.783
p_s	—	—	—	—	—	—
T	0.968	3.472	0.213	0.481	0.782	0.782
P	0.962	<u>3.466</u>	0.213	0.467	<u>0.782</u>	<u>0.783</u>
V	0.959	3.469	0.212	0.471	0.782	0.782
All	0.941	3.458	0.208	<u>0.466</u>	0.783	0.783

Table 5. Performance on integrating various combinations of auxiliary features into FreTS across six cities. Seven configurations of auxiliary feature sets were tested, where results involving service price (p_e) were omitted for JHB, MEL, and SPO. The results demonstrate the usage of different auxiliary features brings different degrees of performance improvement. These findings confirm the value of auxiliary features provided by CHARGED in supporting effective time-series forecasting. * — : Omitted results for cities without p_e .

Usage on Foundation Model. To test the performance of CHARGED in supporting long-term charging volume forecasting, we ran related experiments on foundation models from three model families (Chronos-T5, Moirai-1.1-R, and Moment-1) at small, base, and large scales. Without loss of generality, zero-shot inference was conducted at the site-level for SZH to forecast daily demand for the subsequent month based on the preceding five months of daily-resolution data. Figure 4 depicts the regression fits for each model, with predicted points tightly clustered around the ideal line $y = x$, indicating minimal prediction error and negligible systematic bias. Specifically, the EVS and R^2 values further confirmed the ability of these models to capture virtually all temporal variations and trends. These findings not only demonstrate the excellent zero-shot performance of the foundation model in time series forecasting, but also highlight the critical role of the hourly and citywide datasets provided by CHARGED in enhancing model generalization.

Code availability

CHARGED together with all prepared scripts and tools for data analysis and model evaluation, is publicly available on GitHub at <https://github.com/IntelligentSystemsLab/CHARGED>.

Received: 9 May 2025; Accepted: 8 July 2025;

Published online: 17 July 2025

References

1. Bharathidasan, M. *et al.* A review on electric vehicle: Technologies, energy trading, and cyber security. *Energy Reports* **8**, 9662–9685 (2022).
2. Hsieh, I. Y. L. *et al.* An integrated assessment of emissions, air quality, and public health impacts of China's transition to electric vehicles. *Environmental science & technology* **56**, 6836–6846 (2022).
3. Unterluggauer, T., Rich, J., Andersen, P. B. & Hashemi, S. Electric vehicle charging infrastructure planning for integrated transportation and power distribution networks: A review. *ETransportation* **12**, 100163 (2022).
4. Qureshi, K. N., Alhudaif, A. & Jeon, G. Electric-vehicle energy management and charging scheduling system in sustainable cities and society. *Sustainable Cities and Society* **71**, 102990 (2021).
5. Schoenberg, S., Buse, D. S. & Dressler, F. Siting and sizing charging infrastructure for electric vehicles with coordinated recharging. *IEEE Transactions on Intelligent Vehicles* **8**, 1425–1438 (2022).
6. Ren, H., Zhou, Y., Wen, F. & Liu, Z. Optimal dynamic power allocation for electric vehicles in an extreme fast charging station. *Applied Energy* **349**, 121497 (2023).
7. Su, S. *et al.* Electric Vehicle Charging Guidance Strategy Considering “Traffic Network-Charging Station-Driver” Modeling: A Multiagent Deep Reinforcement Learning-Based Approach. *IEEE Transactions on Transportation Electrification* **10**, 4653–4666 (2023).
8. Kavianipour, M. *et al.* Electric vehicle fast charging infrastructure planning in urban networks considering daily travel and charging behavior. *Transportation Research Part D: Transport and Environment* **93**, 102769 (2021).
9. Wu, Y. *et al.* Hierarchical operation of electric vehicle charging station in smart grid integration applications—An overview. *International Journal of Electrical Power & Energy Systems* **139**, 108005 (2022).
10. Elvas, L. B. & Ferreira, J. C. Intelligent transportation systems for electric vehicles. *Energies* **14**, 5550 (2021).
11. Dijk, D. V., Teräsvirta, T. & Franses, P. H. Smooth transition autoregressive models—a survey of recent developments. *Econometric reviews* **21**, 1–47 (2002).
12. Box, G. E. & Pierce, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association* **65**, 1509–1526 (1970).
13. Yi, K. *et al.* Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems* **36**, 76656–76679 (2023).
14. Cheng, M. *et al.* ConvtimeNet: A deep hierarchical fully convolutional model for multivariate time series analysis. In *Companion Proceedings of the ACM on Web Conference 2025*, 171–180 (2025).
15. Lin, S. *et al.* Segrnn: Segment recurrent neural network for long-term time series forecasting. *arXiv preprint arXiv:2308.11200* (2023).
16. Naghashi, V., Boukadoum, M. & Diallo, A. B. A multiscale model for multivariate time series forecasting. *Scientific Reports* **15**, 1565 (2025).
17. Qu, H. *et al.* ChatEV: Predicting electric vehicle charging demand as natural language processing. *Transportation Research Part D: Transport and Environment* **136**, 104470 (2024).
18. Wang, C. *et al.* Chattime: A unified multimodal time series foundation model bridging numerical and textual data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 12694–12702 (2025).
19. Rashid, M., Elfouly, T. & Chen, N. A comprehensive survey of electric vehicle charging demand forecasting techniques. *IEEE Open Journal of Vehicular Technology* **5**, 1348–1373 (2024).
20. Ma, R. *et al.* Spatial heterogeneity analysis on distribution of intra-city public electric vehicle charging points based on multi-scale geographically weighted regression. *Travel Behaviour and Society* **35**, 100725 (2024).
21. You, L. *et al.* A framework reforming personalized Internet of Things by federated meta-learning. *Nature Communications* **16**, 1–13 (2025).
22. Lu, Y. *et al.* A Retrieval-Augmented Generation Framework for Electric Power Industry Question Answering. In *Proceedings of the 2024 2nd International Conference on Electronics, Computers and Communication Technology*, 95–100 (2024).
23. Yang, J., Rahardja, S. & Fränti, P. Outlier detection: How to threshold outlier scores? In *Proceedings of the international conference on artificial intelligence, information processing and cloud computing*, 1–6 (2019).
24. Panfilo, G. & Arias, F. The coordinated universal time (UTC). *Metrologia* **56**, 042001 (2019).
25. Ester, M., Kriegel, H. P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231 (1996).
26. Guo, Z. *et al.* A City-scale and Harmonized Dataset for Global Electric Vehicle Charging Demand Analysis. *Zenodo* <https://doi.org/10.5281/zenodo.15638530> (2025).
27. Ansari, A. F. *et al.* Chronos: Learning the Language of Time Series. *arXiv preprint arXiv:2403.07815* (2024).
28. Woo, G. *et al.* Unified Training of Universal Time Series Forecasting Transformers. In *Proceedings of the 41st International Conference on Machine Learning*, 53140–53164 (2024).
29. Goswami, M. *et al.* MOMENT: A Family of Open Time-series Foundation Models. In *Proceedings of the 41st International Conference on Machine Learning*, 16115–16152 (2024).
30. Lee, Z. J., Li, T. & Low, S. H. Acn-data: Analysis and applications of an open ev charging dataset. In *Proceedings of the tenth ACM international conference on future energy systems*, 139–149 (2019).

31. Sørensen, Å.L., Lindberg, K. B., Sartori, I. & Andresen, I. Analysis of residential EV energy flexibility potential based on real-world charging reports and smart meter data. *Energy and Buildings* **241**, 110923 (2021).
32. Obusevs, A., Domenico, D. D. & Korba, P. One Year Recordings of Electric Vehicle Charging Fleet. *IEEE Dataport* <https://doi.org/10.21227/fkap-fr63> (2021).
33. Baek, K., Lee, E. & Kim, J. A dataset for multi-faceted analysis of electric vehicle charging transactions. *Scientific Data* **11**, 262 (2024).
34. Li, H. *et al.* UrbanEV: An open benchmark dataset for urban electric vehicle charging demand prediction. *Scientific Data* **12**, 523 (2025).

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China(2023YFB4301900), Research Funds from the Department of Science and Technology of Guangdong Province(2021QN02S161), and the GuangDong Basic and Applied Basic Research Foundation (2023A1515012895).

Author contributions

Linlin You, Rui Zhu, and Chau Yuen designed research. Linlin You, Rui Zhu, and Yan Zhang performed investigation. Zihan Guo, Linlin You, and Chau Yuen collected and preprocessed data. Zihan Guo and Rui Zhu worked together to conceive and conduct the experiments. Linlin You, Rui Zhu and Yan Zhang analyzed and discussed the experimental results. All authors contributed to the writing and revising of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025