www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

A high-resolution electric vehicle charging transaction dataset with multidimensional features in China

Yuanshi Zhang[1,2], Tongxin Xu[1,2], Tao Chen[1,2], Qinran Hu[1,2] ✉, Hongrui Chen[1,2], Xukun Hu[3] & Zilv Jiang[3]

The charging transaction data of electric vehicle (EV) users is crucial for studying charging market dynamics and formulating effective policies. However, due to factors such as the privacy of EV users and the complex coupling relationships between charging dealers, existing EV charging transaction datasets are plagued by issues such as incompleteness, significant bias, and a lack of real-time information. To address these issues, a real-time charging transaction dataset has been created, comprising 441,077 charging transactions collected from 13 charging stations in China over a 2-year period. The dataset includes detailed data of EV user charging transaction time, price and charging status, as well as the charging termination reasons and weather data for each charging session. This dataset offers references for identifying EV user behaviors and extracting charging fault factors from multiple aspects, supporting research applications in EV charging facility planning, EV charging and discharging management, and charging economic evaluation.

Background & Summary

Greenhouse gas emissions and the global energy crisis are driving emission reduction plans, electric vehicles (EVs), as a low-carbon, convenient, and economical mode of transportation, are one of the most promising solutions for achieving carbon reduction goals in the transportation sector[1]. According to the International Energy Agency (IEA), Global EV sales rise to 14 million units in 2023, representing 18% of all vehicle sales, up 14% compared to the preceding year[2]. The increase in the share of EV has led to research into the proper charging and discharging of EVs being a current hotspot. Current research on EV charging focuses on key areas such as charging station layout optimization[3], charging behavior analysis[4], charging load prediction[5], network management[6], etc., which is highly dependent on the support of charging transaction datasets. Through data mining, machine learning and simulation modelling, charging transaction datasets can not only reveal patterns and predict demand, but also significantly improve the precision of analysis and accuracy of models, thus optimizing management and decision-making efficiency.

Many datasets of EV charging transaction records have been published. Examples of publicly available are as follows. Electric Vehicle Charging Dataset[7] comprises extensive information on EV charging transactions, employing Conditional Tabular Generative Adversative Networks (CTGAN) and Kernel Density Estimation (KDE) to simulate owner behavior. EV Charging Dataset by Leeds City Council[8] offers detailed charging transaction data for a specific location in Leeds from 2014 to 2021, with a temporal resolution of one minute. High-resolution EV Charging Data in the Workplace[9] provides a comprehensive workplace charging dataset with detailed user information, such as vehicle type and commute distance, offering valuable insights for evaluating charging behavior from the user's perspective. Another high-quality dataset[10] has been published covering a broader category of charging station locations across South Korea over one year. In addition, there are a number of other typical charging transaction records such as ElaadNL dataset[11], ST-EVCDP dataset[12] and ACN Data[13]. Datasets in this domain often present certain limitations. For instance, some datasets cover only a brief time period or include a limited number of transaction sessions[8], while several datasets focus solely on transaction data from a single site[9], reducing their generalizability. Moreover, simulation data[7], which is often generated under specific assumptions, can be overly reliant on these assumptions and challenging to validate. Additionally,

[1]School of Electrical Engineering, Southeast University, Nanjing, 210096, China. [2]Jiangsu Provincial Key Laboratory of Smart Grid Technology and Equipment, Southeast University, Nanjing, China. [3]State Grid Zhejiang Electric Power Co., Ltd. Tongxiang Power Supply Company, Jiaxing, China. ✉e-mail: qhu@seu.edu.cn

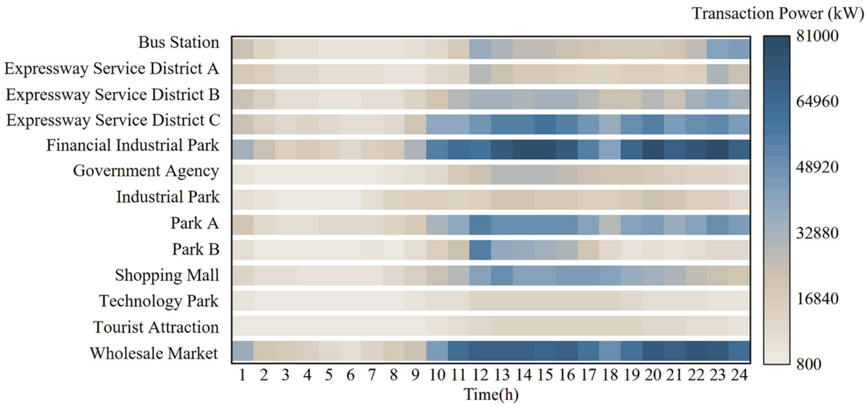

Fig. 1 Distribution of the total number of charges across the 13 charging points within a day.

	Single Phase	Three-Phase, Four-Wire
Communication Interface	4G	RS485
Accuracy Class	Class 1.0	
Voltage Line Apparent/Active Power Consumption	<10VA/2.0W	
Current Line Apparent Power Consumption	<2.5VA	
Starting Current	≤0.004Ib (for Class 1.0), ≤0.005Ib (for Class 2.0)	
Operating Temperature Range	−25 °C ~ +55 °C	
Clock Error	≤0.5s/d (25 °C)	
Relative Humidity	25%～95%	
Storage Temperature	−40 °C ~ +70 °C	

Table 1. Parameters of data collection equipment.

most datasets emphasize the EV charging transaction process but fail to account for critical factors such as the willingness of EV users to charge (e.g., reasons for ending a transaction) and external travel influences, including weather conditions.

To fill this gap, this paper generated a high-resolution, large-scale dataset encompassing an extensive range of spatial dimensions and enriched informational depth to ensure data representativeness and comprehensiveness. Specifically, the dataset comprises information on charging transactions across urban and suburban stations within a Chinese city, including user identification numbers (user IDs), precise charging start and end times (to the second), electricity volumes transacted, associated costs (electricity and service), reasons for termination and daily weather conditions. The dataset is organized by charging station location, with derived variables generated according to transaction termination reasons, thereby facilitating streamlined data processing. The dataset offers valuable applications in the following scenarios:

1) Adjustment of the layout of the charging stacks according to the charging behavior of the charging users in order to improve the user experience[14,15];
2) Evaluate the effectiveness of EV support policies and develop more accurate market strategies[16,17];
3) Investigate the impact of weather factors on battery performance and user charging habits, and provide data references for EV charging load prediction and simulation of load profiles taking into account weather[18,19];
4) Analyze the failure rate and causes of charging stacks to assess equipment stability and system reliability, and provide guidance for equipment operation, maintenance and troubleshooting[20].

Methods

Data collection. The charging transaction dataset was obtained through collaborative agreements with State Grid Zhejiang Electric Power Co., Ltd. (Tongxiang Power Supply Company) and the Jiangsu Provincial Key Laboratory of Smart Grid Technology and Equipment. The raw operational data was desensitized and anonymized to remove all identifying information, with consent of the participants collected by the industry partner. The providers confirmed compliance with internal data policies and ethical review exemptions, as no human subjects were involved. The temporal distribution of total charging at the 13 charging stations is shown in Fig. 1. Charging is initiated by users either by scanning a QR code at the station, logging into their service platform account, and selecting the station, or through identity verification on the platform. The platform supports intelligent vehicle networking capabilities, enabling real-time monitoring and management of charging equipment, dynamic load balancing, automatic billing, and efficient operational management. During charging, sensors

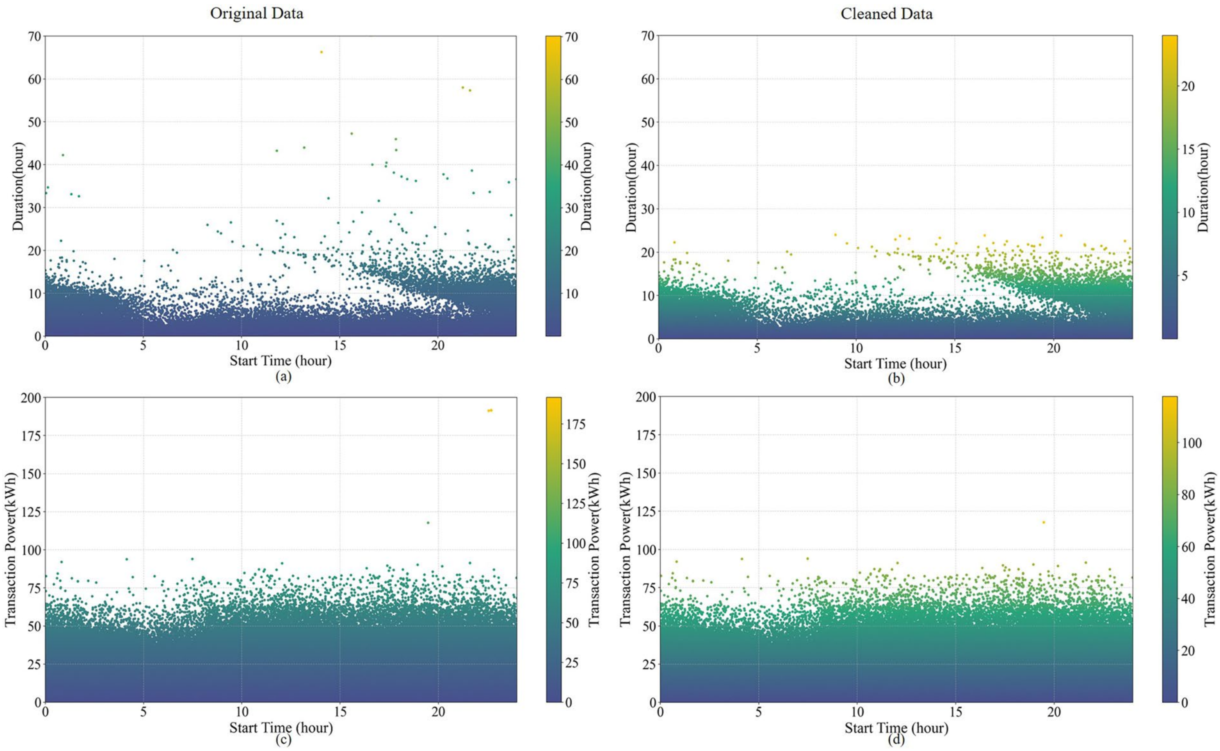

Fig. 2 Impact of outlier treatment on data distributions: (**a,c**) Original data; (**b,d**) Cleaned data.

End cause derived variables	Concrete description
Is_full_charge	The pre-charge is used up or the electric car is full.
Is_user_stop	The user cancels remotely over the network or spontaneously at the charging station.
Is_nonsystem_fault	Failure of lines or components in the main or control circuit, such as AC and DC failures, abnormal voltage and current at input and output terminals, BMS failure, auxiliary power supply failure, insulation monitoring failure, etc.
Is_system_issue	Anomalies in startup, temperature, system address configuration, communication, or human-machine interaction malfunctions.
Is_EV_fault	EV communication failures or anomalies in battery parameters and insulation status during charging.
Is_other_fault	Failures caused by improper user operation and other manufacturer-defined failures.
Is_abnormal	A binary variable that indicates whether the session ended abnormally.

Table 2. Concrete description of end cause derived variables.

continuously monitor the operational status of the charging stack—including metrics such as charging power, power curve, battery voltage, and current—collecting data in real time. The device follows the IEC 60870-5-104 protocol (IEC 104) and communicates via Bluetooth and 4G dual-module set, lower supports 5-minute data reporting and ensures standardized data exchange between the vehicle and the charger. The detailed information of the data collection devices is displayed in Table 1. Transaction data is transmitted through a cellular network to a cloud server, subsequently relayed to a back-end management system. To ensure data integrity and security, SSL/TLS encryption is employed during transmission to prevent potential eavesdropping or unauthorized access. A timeout of five seconds is triggered if the charger and Battery Management System (BMS) fail to receive the correct or isolation messages within the specified timeframe. If a timeout occurs, the BMS or charger sends an error message, initiating an error-handling state. Error processing involves executing distinct procedures tailored to specific error categories[20]. Meteorological data were sourced from the China Meteorological Data Service Centre (CMDSC), with daily station observations interpolated to a 0.1° × 0.1° grid. Zonal averaging was applied to derive district-level weather parameters for Jiaxing's three administrative regions. This dataset uses day-by-day interpolated meteorological data to assist the analysis and does not disclose raw station-level data. This disclosure is in accordance with Measures for the Management of Meteorological Data Sharing.

Data pre-processing.

1) Anonymization: In order to mitigate the risk of data leakage and misuse, as well as to support the long-term storage and utilization of data, we have removed all the user-specific identifying information, including mobile phone numbers, VINs, vehicle registration numbers, and transaction slip numbers. User IDs are then desensitized using random mapping.

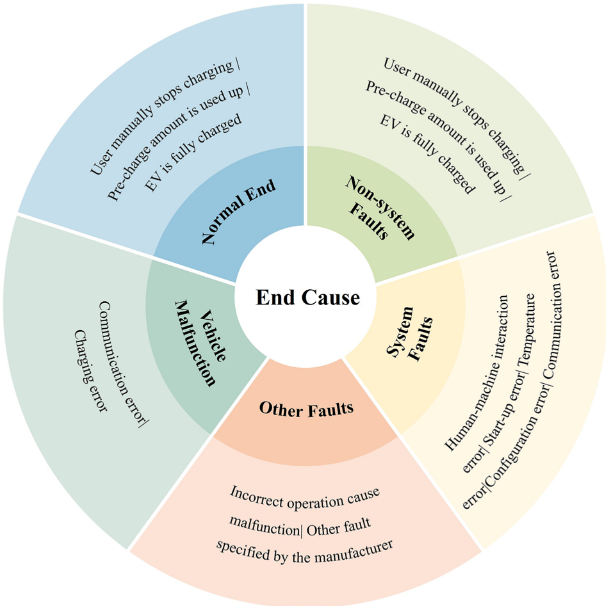

Fig. 3 Classification criteria for variables derived from end cause.

	Is_user_stop			Is_full_charge		
Median charging duration(min)	30.9			46.7		
Charging capacity(kWh)	**25%**	**Median**	**75%**	**25%**	**Median**	**75%**
	6.5	14.3	23.4	11.8	19.7	29.3

Table 3. Statistics on normal end of charge.

Column	Concrete description
User ID	A unique identifier is assigned to each user and allows the system to track and manage user-specific data.
Charging Post ID	A unique identifier for each charging post or station and differentiates between various posts in a network.
Order Creation Time	The timestamp when the system creates the charging order, marking the start of the transaction process.
Transaction Power (kWh)	The amount of electricity consumed during the session, measured in kilowatt-hours (kWh), indicating how much energy was used.
Electricity Cost (Yuan)	The cost of the electricity consumed, calculated based on the energy used (in kWh) and the rate per kWh.
Service Charge (Yuan)	The additional charge for using the charging service, covering maintenance, operational, or platform fees.
Transaction Amount (Yuan)	The total amount the user pays for the charging session, including both electricity costs and service charges.
Actual Payment (Yuan)	The portion paid by the user after deducting coupons, discounts or other allowances.
Start Time	The time when the user begins charging the vehicle, marking the start of the charging process.
End Time	The time when the charging session completes, indicating the end of the charging process.
Payment Time	The timestamp when the user completes the payment for the charging session, typically after the session ends.
End Cause Derived Variables	Variables describing the reason for ending the session, such as a full battery, user cancellation, or system failure.
Temperature (°C)	The average temperature at each site over a 24-hour period.
Relative Humidity (%)	The ratio between absolute humidity and maximum humidity, showing the degree of saturation of water vapor.
Precipitation(mm)	The amount of precipitation per unit area that falls on the ground, on a horizontal surface, over a given period of time.
Location Information	Information identifying the location category of the charging point.

Table 4. Detailed explanation of the dataset.

2) Handling Missing Data: Transaction sessions that were not fully recorded due to network outages etc. have been removed, which accounted for 0.022% of the complete dataset. Billing transaction data that is not fully recorded due to factors such as network interruptions or equipment failures is excluded from all analyses except outage analyses to prevent incomplete information from negatively impacting the analysis results.
3) Handling Abnormal Data: For charging sessions where charging ends normally, we removed outlier sessions such as charging duration is too long, charging amount exceeds the rated battery capacity and does not meet the logic of charging charges that take into account incentives (e.g., Actual Payment is greater than the sum of the Electricity and Service Charge). Figure 2 shows that the data treated with outliers did

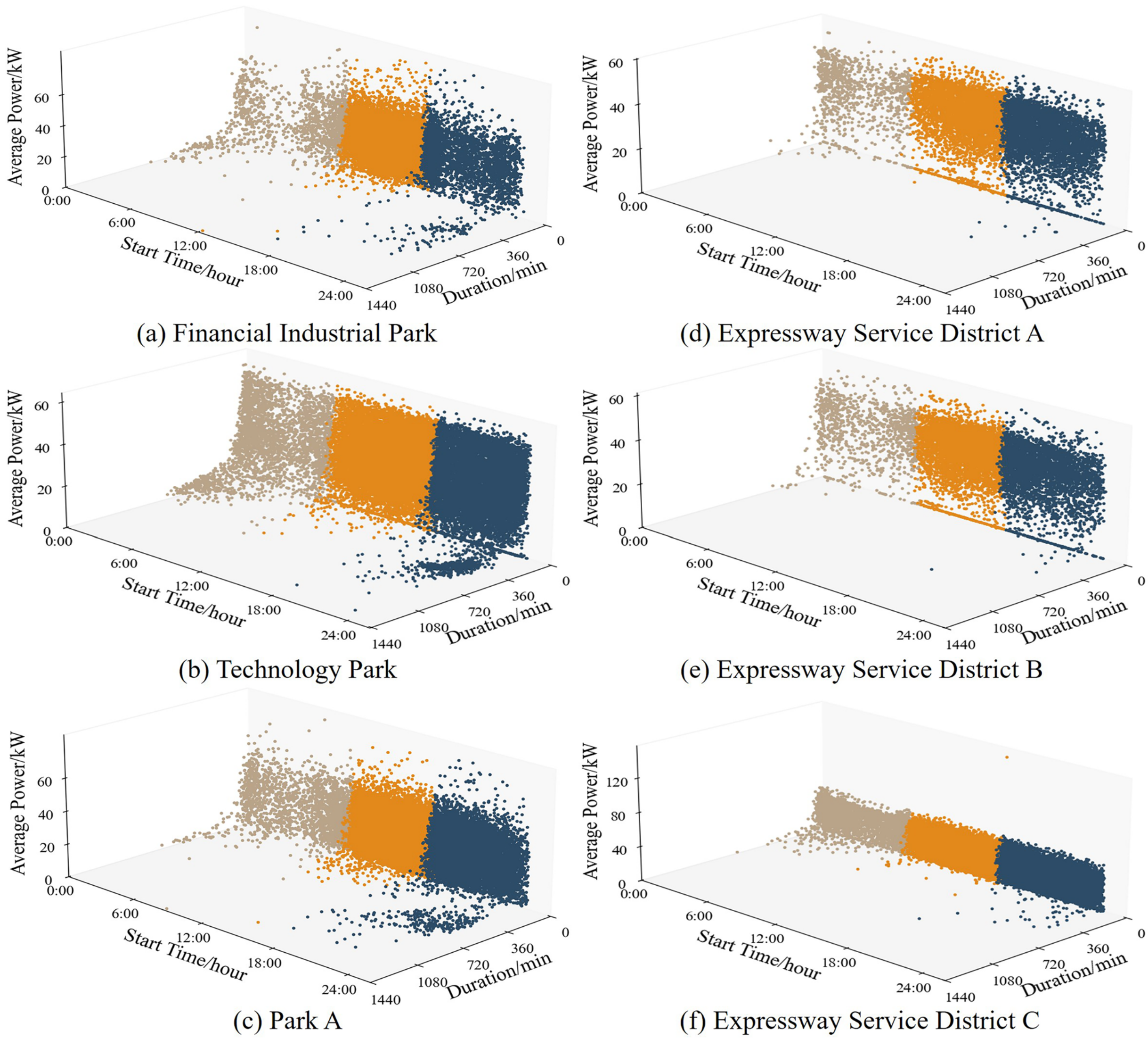

Fig. 4 Cluster analysis results of charging transactions in different areas (e.g. urban, remote).

not significantly modify the statistical information of the dataset. The K-S test verified that the cleaned data were not significantly different from the original data distribution in the key indicators of charging volume (D = 0.00006), cost (D = 0.00006), and charging start time (D = 0.00004).

4) The End of Charge Reason field is derived from the Charging Pile Communication Protocol GB/T 27930-2015 standard message, and this textual information is converted into binary-derived variable as shown in Table 2. Derived variables are classified by reference to the national EV charging facility fault classification criteria. The categories of normal end or fault causes included in the different end causes are shown in Fig. 3. The accuracy of the generation of derived variables can be reacted by comparing the charging normal end statistics in Table 3. The median charging amount where the user stops charging is significantly lower than those of the fully charged records, and the median charging duration is significantly different from that of the fully charged records.

Data Records

All the data presented in this paper have been compiled into a complete dataset, which is available for download from the Figshare[21].

This dataset encompasses a total of 441,077 transaction tasks over a two-year period, collected from 13 charging stations across three districts of Jiaxing, a prefecture-level city in China. The dataset includes charging transaction details such as user IDs, charging start and end times (to the second), amount of electricity transacted, electricity tariffs, service fees, reasons for charging termination, weather conditions and other charging transaction information. And detailed explanation of the dataset can be found in Table 4. The data provided is raw and has undergone verification. Users can apply various data interpolation methods to process the dataset as needed, using the provided scripts.

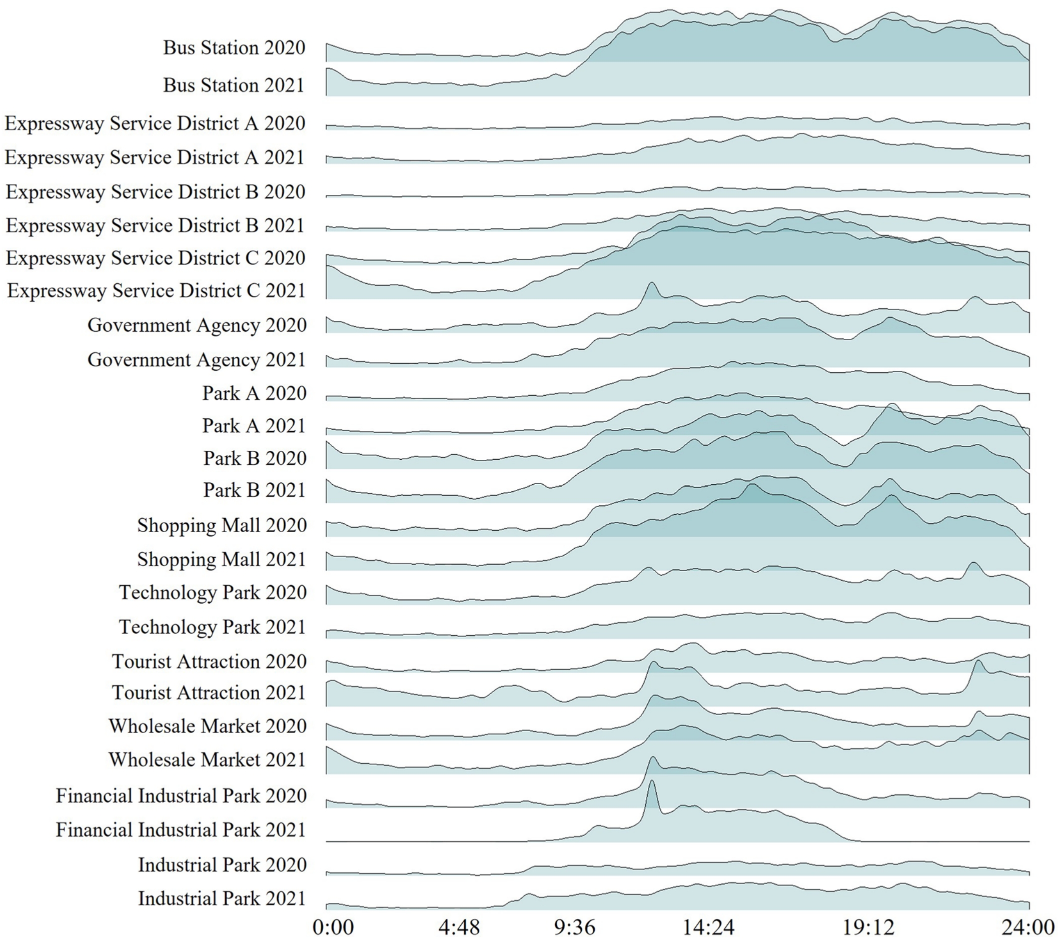

Fig. 5 Temporal distribution of user charging frequency at different sites over two years.

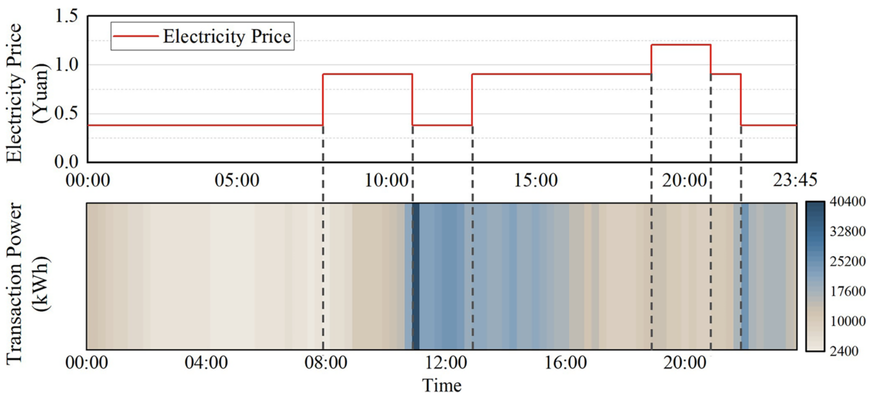

Fig. 6 The relationship between load demand (total electricity traded) and time periods under TOU Price in Nanhu. (**a**) Electricity price; (**b**) Transaction power.

Period	Period type	Electricity Price(¥ /kWh)
19-21	Sharp Pricing	1.2064
8-11,13-19,21-22	Peak Pricing	0.9014
11-13,22-8	Off-peak Pricing	0.3784

Table 5. Local TOU Price (Yuan/kWh) for Different Time Periods[24].

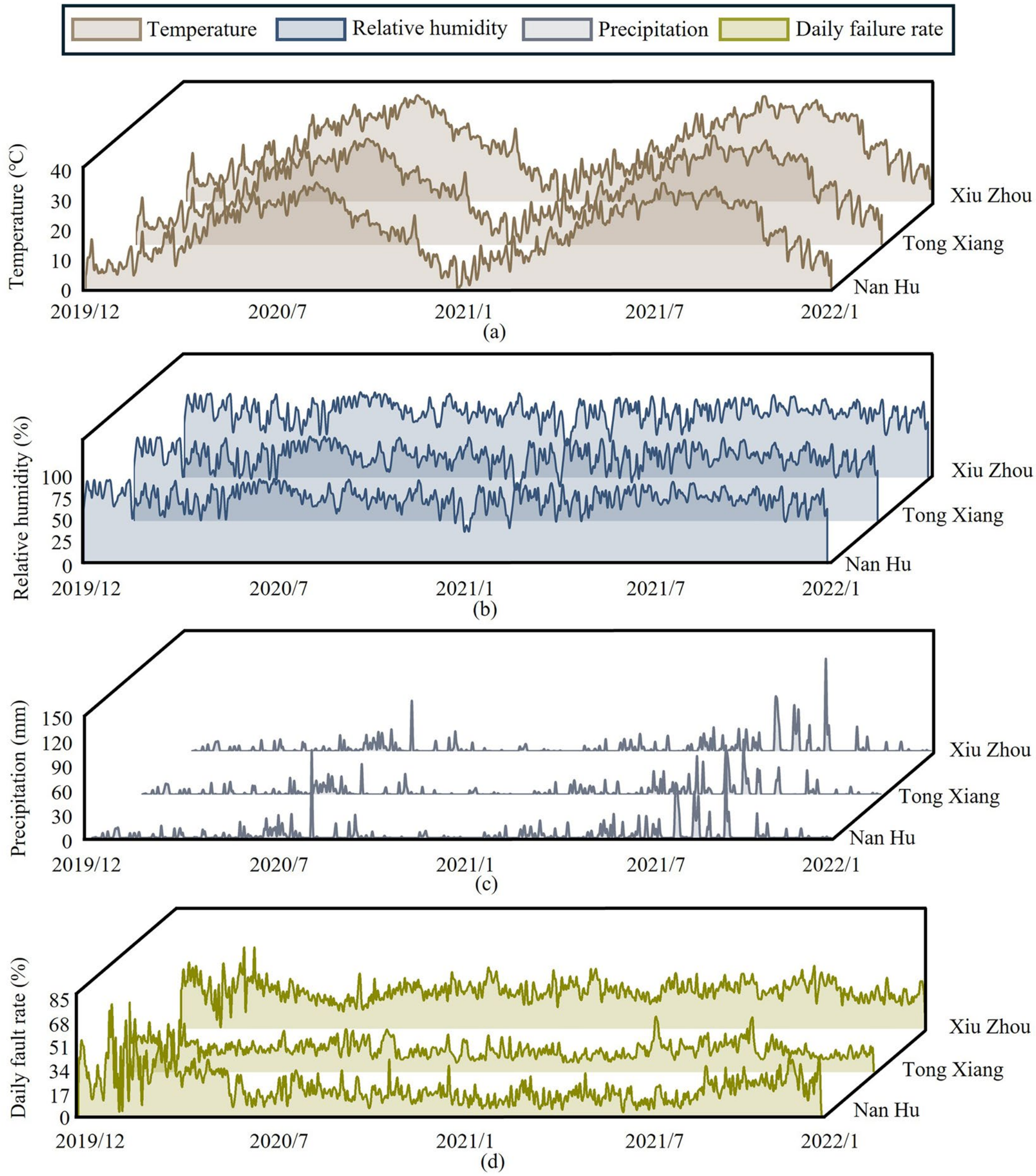

Fig. 7 Relationship between failure rates and weather conditions over two years. (**a**) Temperature; (**b**) Relative Humidity; (**c**) Precipitation; (**d**) Daily fault rate.

Variable Name	P Value	Mean of Lockdown Period	Mean of Recovery Period
Transaction Power/kWh	5.2027e-1	16.2059	16.3582
Charging Duration/h	3.9872e-8	0.9521	0.9305
Is_user_stop	7.0776e-13	0.3119	0.3414
Is_abnormal	3.8724e-64	0.2504	0.1902

Table 6. Comparison of charging behavior during the lockdown and recovery period.

Technical Validation

The dataset comprises real-time information transmitted from local charging stations, ensuring both its authenticity and reliability. Additionally, the charging behavior and statistical patterns of the representations are analyzed in conjunction with the spatio-temporal dimensional characteristics of the dataset to confirm the accuracy and relevance of the data.

Analysis of charging habits. All available charging data in urban areas and remote locations has been visualized and analyzed using the K-means algorithm, as shown in Fig. 4. There is a significant difference in the

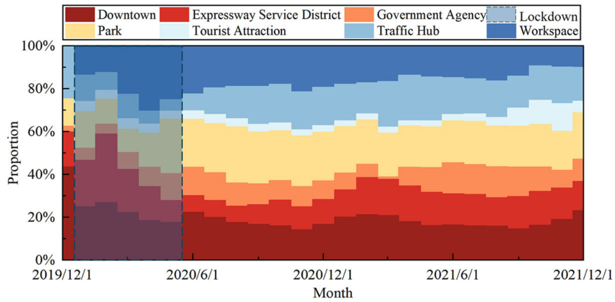

Fig. 8 Analysis of spatio-temporal heterogeneity in charging behavior.

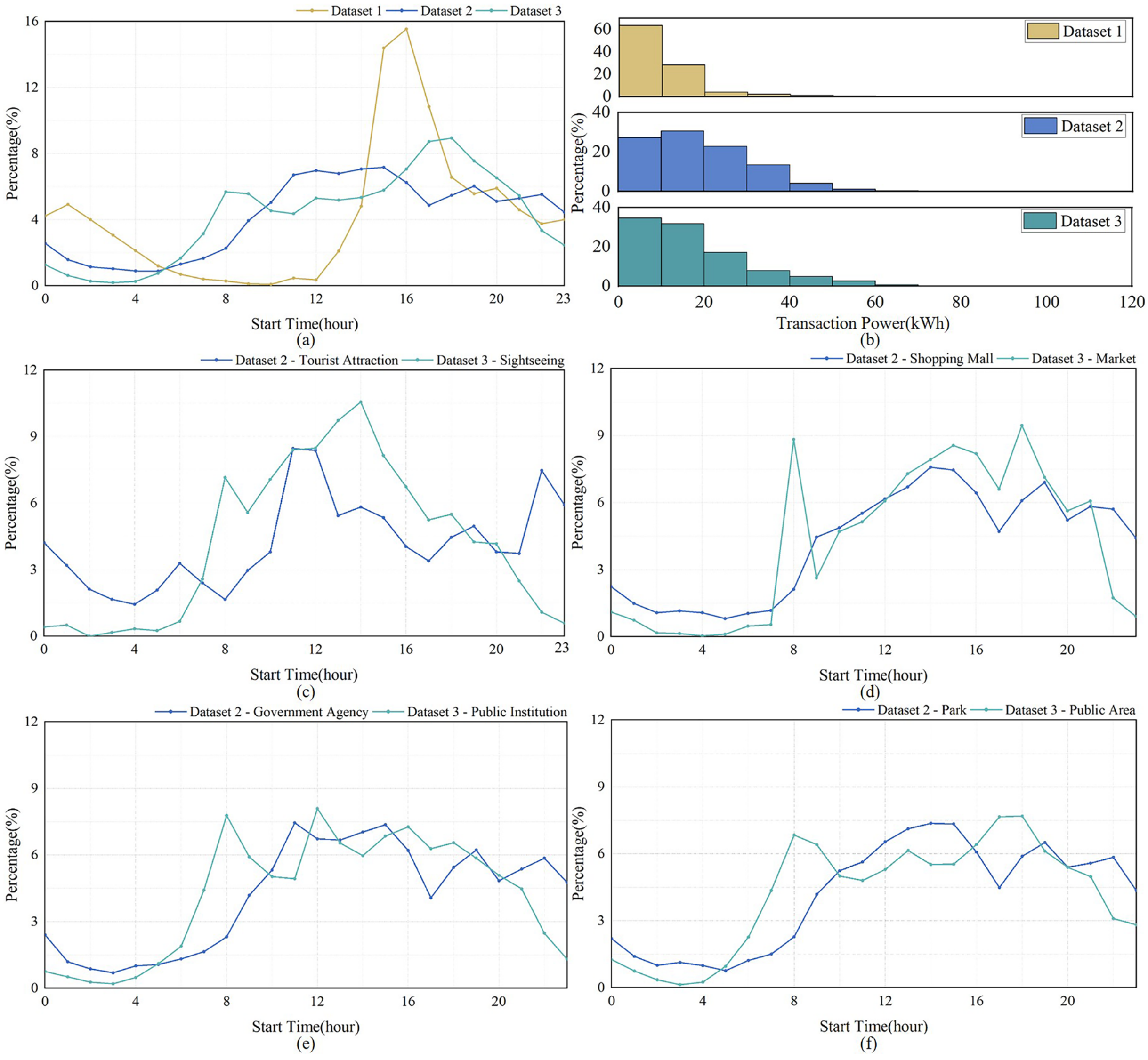

Fig. 9 Comparative analysis of EV charging behaviors across datasets:(**a**) Session duration distributions, (**b**) Transaction energy distributions, (**c–f**) Location-specific duration patterns. (1: ACN Data; 2: Proposed Dataset; 3: Reference Dataset[10]).

temporal distribution of fast and slow chargers in urban areas, while there are fewer slow chargers in expressway service areas. The temporal distribution of charging frequency at each site over the two years is presented in Fig. 5. The charging connection time patterns for the same location exhibit strong consistency, thus confirming the validity and reliability. The relationship between time-of-use (TOU) price in Jiaxing (as is shown in Table 5) and load demand is shown in Fig. 6.

Dataset	Mean	Minimum	Maximum	25%	50%	75%
ACN Data	9.25	0.50	77.70	3.40	6.72	12.95
Proposed Dataset	18.91	0	117.74	9.05	17.15	26.78
Reference dataset in [10]	17.44	0.01	97.00	7.53	14.10	23.20

Table 7. Statistics on energy transaction volumes in different datasets.

Failure rate analysis. The distribution of day-by-day meteorological data and failure rates for the three zones over two years is shown in Fig. 7. By comparing the four subfigures, it is clear that the daily failure rate is relatively higher around September each year, i.e., when the precipitation and temperature are relatively higher.

Impact of COVID-19. The timing of this dataset covers the entire period of the COVID-19 from the outbreak to the resumption of work and normal life. In order to assess potential bias, we conducted subgroup analyses, counting charging duration, charging volume, user probability of stopping charging, and order anomaly probability during the blockade period (first quarter of 2020) and the post-recovery phase (second to fourth quarter of 2021). Also, Mann-Whitney U and Fisher's Exact were used to test for significant differences as is shown in Table 6. There is no significant difference in the amount of charging from one period to the next representing some stability in the data overall. At the same time, the differences in charging time, end of user charging and failure rate at different stages also have interpretability due to the rebound in demand for long-distance travel, which may be related to the change in users' travel plans and the resumption of work by the operation and maintenance team. Figure 8 shows the regional heterogeneity caused by COVID-19.

Generalizability analysis. China's charging facilities are concentrated in economically developed areas such as the Yangtze River Delta and Pearl River Delta, similar to the distribution patterns in Europe (dominated by core markets such as the Netherlands and Germany[22]) and North America (dominated by the east and west coasts of the US[23]). Cross-dataset validation against two open academic benchmarks reveals fundamental consistency in behavioral patterns, particularly in charging session timing (Fig. 9a,b) and energy transaction distributions (Table 7). Functional-equivalent stations exhibit synchronized daily charging profiles aligning with established patterns in[10] (Fig. 9c–f), confirming operational regularity across datasets.

Code availability

The analysis code was written using Python 3.10, and the visual plots were realized through Origin 2021. The purpose of each script is mentioned in the comments, and the code was uploaded to Figshare along with the dataset[21].

Received: 15 January 2025; Accepted: 9 April 2025;
Published online: 16 April 2025

References

1. Fang, T., Vairin, C., von Jouanne, A., Agamloh, E. & Yokochi, A. Review of Fuel-Cell Electric Vehicles. *Energies* **17**, 2160 (2024).
2. International Energy Agency, Global EV Outlook https://www.iea.org/reports/global-ev-outlook-2024 (2024).
3. Su, H., Duan, L., Luo, Q. & Song B. Layout Optimization of Smart Charging Piles on Highways Based on Nonlinear Integer Programming, *2022 International Symposium on Intelligent Robotics and Systems (ISoIRS)*, 181–186 (2022).
4. Wang, Y., Yao, E. & Pan, L. Electric vehicle drivers' charging behavior analysis considering heterogeneity and satisfaction. *Journal of Cleaner Production* **286**, 124982 (2021).
5. Huang, X., Wu, D. & Boulet, B. MetaProbformer for Charging Load Probabilistic Forecasting of Electric Vehicle Charging Stations. *IEEE Transactions on Intelligent Transportation Systems* **24**, 10445–10455 (2023).
6. Bian, P. *et al*. Optimal Regulation Strategy for Electric Vehicle Load Clusters based on Demand Response, *2023 IEEE International Conference on Power Science and Technology (ICPST)*, 606–610 (2023).
7. Gholizadeh, N., Electric Vehicle Charging Dataset, *Mendeley Data* https://doi.org/10.17632/5zrtmp7gwd.1 (2024).
8. Leeds City Council, *Electric vehicle chargepoints operated by Leeds City Council*. https://www.data.gov.uk/dataset/5bb5c097-0e2f-42a3-8aee-1c0189a39082/electric-vehicle-chargepoints-operated-by-leeds-city-council (2021).
9. Asensio, O. I., Lawson, M. C. & Apablaza, C. Z. Metadata record for: Electric vehicle charging stations in the workplace with high-resolution data from casual and habitual users. *figshare*. https://doi.org/10.6084/m9.figshare.14357123.v1 (2021).
10. Baek, K., Lee, E. & Kim, J. A dataset for multi-faceted analysis of electric vehicle charging transactions. *figshare*. https://doi.org/10.6084/m9.figshare.22495141.v1 (2023).
11. Elaad, N. L. ElaadNL Open Datasets for Electric Mobility Research. https://www.elaad.nl (2020).
12. Kuang, H. *et al*. Unraveling the effect of electricity price on electric vehicle charging behavior: A case study in Shenzhen, China. *Sustainable Cities and Society* **115**, 105835 (2024).
13. Lee, Z. J., Li, T. & Low, S. H. ACN-Data: Analysis and Applications of an Open EV Charging Dataset. *Tenth ACM International Conference on Future Energy Systems*, 139–149 (2019).
14. Zhang, Y. *et al*. Efficient Deployment of Electric Vehicle Charging Infrastructure: Simultaneous Optimization of Charging Station Placement and Charging Pile Assignment. *IEEE Transactions on Intelligent Transportation Systems* **22**, 6654–6659 (2021).
15. Liang, Z., Qian, T., Korkali, M., Glatt, R. & Hu, Q. A Vehicle-to-Grid planning framework incorporating electric vehicle user equilibrium and distribution network flexibility enhancement. *Applied Energy* **376**, 124231 (2024).
16. Limmer, S. & Rodemann, T. Peak load reduction through dynamic pricing for electric vehicle charging. *International Journal of Electrical Power & Energy Systems* **113**, 117–128 (2019).
17. Gupta, A. K. & Bhatnagar, M. R. A Comprehensive Pricing-Based Scheme for Charging of Electric Vehicles. *IEEE Systems Journal* **17**, 3492–3502 (2023).
18. Yan, J. *et al*. EV charging load simulation and forecasting considering traffic jam and weather to support the integration of renewables an EVs. *Renewable energy* **159**, 623–641 (2020).
19. Arias, M. B. & Bae, S. Electric vehicle charging demand forecasting model based on big data technologies. *Applied Energy* **183**, 327–339 (2016).

20. Wang, W. *et al*. Exploring best-matched embedding model and classifier for charging-pile fault diagnosis. *Cybersecurity* **6**, 7 (2023).
21. Zhang, Y. *et al*. A high-resolution electric vehicle charging transaction dataset with multidimensional features in China. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.28182251.v3 (2025).
22. Electric cars: Half of all chargers in EU concentrated in just two countries. https://www.acea.auto/press-release/electric-cars-half-of-all-chargers-in-eu-concentrated-in-just-two-countries/ (2022).
23. Electric Vehicle Charging Station Locations. https://afdc.energy.gov/fuels/electricity-locations#/find/nearest?fuel=ELEC.
24. Jiaxing Municipal People's Government. https://www.jiaxing.gov.cn/art/2019/9/25/art_1685296_38320694.html (2019).

Acknowledgements
This work is supported by the Jiangsu Provincial Key Laboratory Project of Smart Grid Technology and Equipment.

Author contributions
Conceptualization, Y.Z., T.X.; Methodology, T.C., Q.H.; Visualization & writing, Y.Z., T.X., T.C., Q.H.; Investigation, H.C. X.H., Z.J.; Project Administration & funding acquisition, X.H., Z.J.; Supervision, T.C., X.H., Z.J.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Q.H.

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.