www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

A high-resolution dataset on electric passenger vehicle characteristics in China and the European Union

Lang Mai[1,8], Ming Liu[1,2,8], Han Hao[1,3] ✉, Xin Sun[4], Fanran Meng[5], Yong Geng[6], Zia Wadud[7], James E. Tate[7], Zhenyu Dong[1], Haoyang Li[1], Jingxuan Geng[1], Hao Dou[1], Yunfeng Deng[1], Fanlong Bai[1], Zongwei Liu[1] & Fuquan Zhao[1]

China and the EU are the world's largest Electric Vehicle (EV) markets, making it crucial to understand their electrification progress for global insights. However, previous assessments of regional EV markets often provide broad EV market characteristic estimations, but neglect critical spatial and segmental heterogeneity, thereby limiting research and policy precision. To fill such a knowledge gap, this study proposes a multi-dataset fusion approach that enables the characterization of passenger vehicle electrification progress in both China and the EU at highly resolved spatial, segmental, and powertrain levels for the year 2023. The dataset includes EV sales, market penetration, battery chemistry mix, and sales-weighted average battery capacity for all wheelbase-defined segments across 31 provinces and municipalities in China, as well as the EU27, Iceland, and Norway. It characterizes the current state of passenger vehicle electrification in China and the EU and supports further research on critical material demand estimation, decarbonization performance assessment, and related topics.

Background & Summary

The global energy transition and climate change mitigation efforts have driven the rapid development of Electric Vehicles (EVs)[1–4]. In 2023, the sales of the global electric passenger vehicles reached 13.8 million units, increasing by more than 35% compared to 2022[5]. Also, the global electric passenger vehicle penetration rate achieved 18%[6], which is expected to exceed 80% by 2040 under net-zero emission targets[7]. Behind such rapid electrification is the frequent introduction of new EV models into the market, with nearly 590 models available globally in 2023, surpassing any of the previous years[5]. This makes the EV market characteristics, including battery capacity, battery chemistry mix, among others, undergo rapid changes[8], imposing complex and fast-changing impacts on energy, resources, and the environment[9–12].

China and Europe are the top two EV markets in the world, together accounting for over 80% of the global EV sales in 2023[13]. Since 2018, available EV models in China and European countries have doubled and tripled, respectively[14], making them the frontiers in EV market characteristics evolution. It is, therefore, imperative to understand their electrification progress for global insights. However, high complexity lies in the characteristics of these two large markets, considering numerous sub-regions with diverse EV development levels in them, varied vehicle segments with distinct technological features, and multiple powertrain systems with different attributes. This introduces large uncertainties to research and impedes valuable, customized policy implications,

[1]School of Vehicle and Mobility, Tsinghua University, Beijing, 100084, China. [2]Energy Policy Research Group, Cambridge Judge Business School, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, UK. [3]Tsinghua-Rio Tinto Joint Research Center for Resources Energy and Sustainable Development, Tsinghua University, Beijing, 100084, China. [4]Integrated Research on Energy, Environment and Society (IREES), Energy and Sustainability Research Institute Groningen (ESRIG), University of Groningen, Groningen, 9747AG, the Netherlands. [5]Department of Chemical & Biological Engineering, Faculty of Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK. [6]School of International and Public Affairs, Shanghai Jiao Tong University, Shanghai, 200240, China. [7]Institute for Transport Studies, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK. [8]These authors contributed equally: Lang Mai, Ming Liu. ✉e-mail: hao@tsinghua.edu.cn

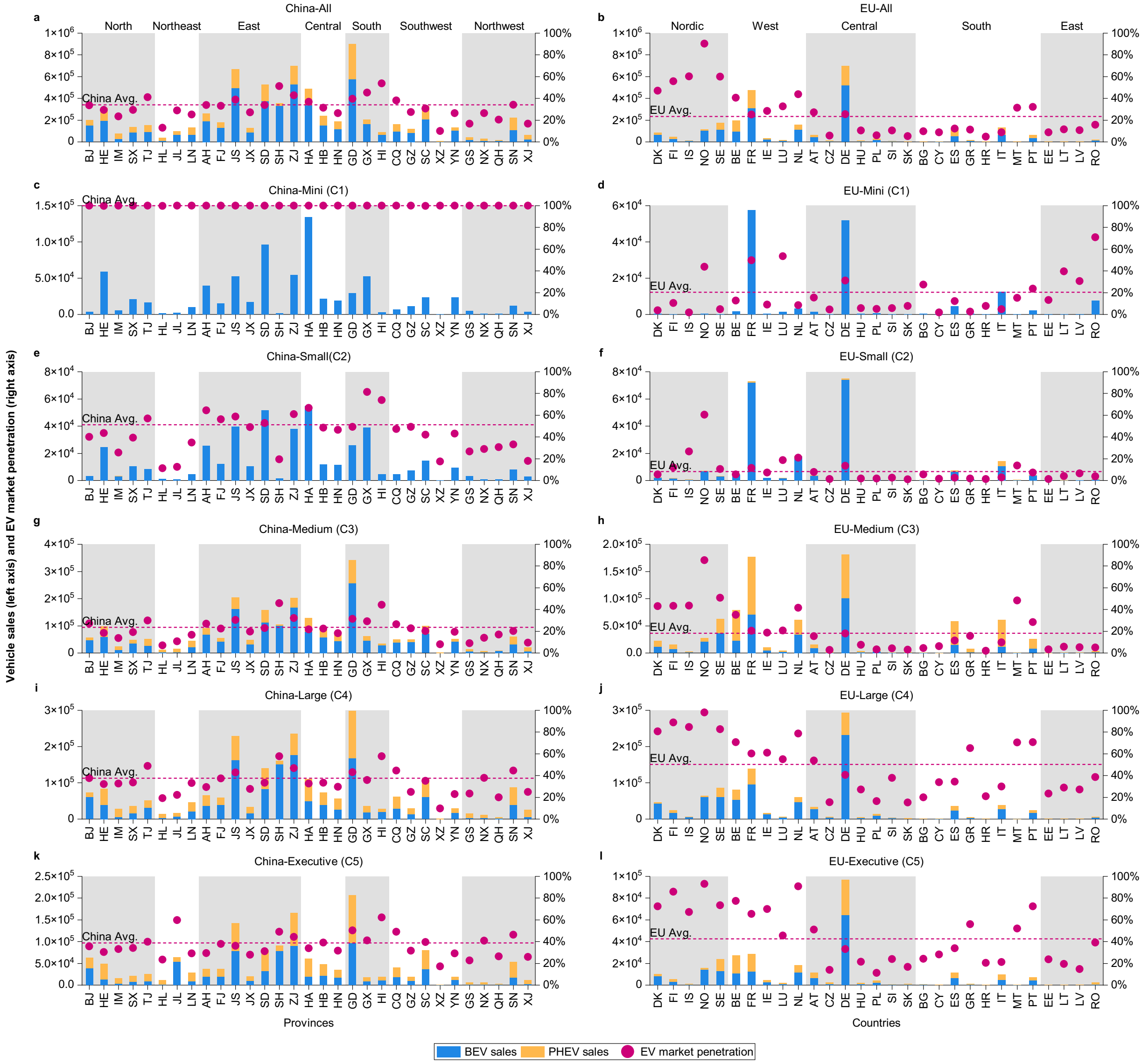

Fig. 1 EV sales and market penetration in China and its provinces (**a,c,e,g,i,k**) and the EU and its member states (**b,d,f,h,j,l**). Chinese provinces and EU countries are divided into 7 and 5 groups by geographical proximity. Dashed lines represent the EV market penetration at the overall level for China or the EU. The full names of the abbreviations are shown in Table 1. EV, Electric Vehicle; BEV, Battery Electric Vehicle; PHEV, Plug-in Hybrid Electric Vehicle; Avg., Average.

thereby necessitating a comprehensive characterization of EV fleets in China and Europe at the finest possible level of granularity.

Despite such an urgent need, existing datasets lack high-resolution characterization that can effectively serve this purpose. EV sales and market penetration assessed in most previous studies are generally resolved at the country level[6], which cannot reveal the internal heterogeneity within vast regions such as China. Also, regional- or segmental-level descriptions of either battery capacity or battery chemistry mix rarely exist, hindering accurate environmental and resource impact assessments. Although some data-driven studies have provided high-resolution characterizations from spatial or segmental perspectives for their specific research purposes[15–17], these typically reflect only a subset of aggregated characteristics and do not offer complete, unaggregated, high-granularity foundational data. Overall, holistic and systematic high-resolution characterizations of the EV markets are still rare.

To address such a knowledge gap, this study adopts a bottom-up multi-dataset fusion method to provide detailed insights into major EV market characteristics: EV market penetration (example of results in Fig. 1), segment mix, battery chemistry mix, and sales-weighted average battery capacity across five passenger vehicle segments spanning from China's provinces to EU countries (EU27 plus Iceland and Norway) in 2023. This dataset can support a range of researches and applications, including:

abbreviation	AH	BJ	CQ	FJ	GD	GS
full name	Anhui	Beijing	Chongqing	Fujian	Guangdong	Gansu
abbreviation	GX	GZ	HA	HB	HE	HI
full name	Guangxi	Guizhou	Henan	Hubei	Hebei	Hainan
abbreviation	HL	HN	IM	JL	JS	JX
full name	Heilongjiang	Hunan	Inner Mongolia	Jilin	Jiangsu	Jiangxi
abbreviation	LN	NX	QH	SC	SD	SH
full name	Liaoning	Ningxia	Qinghai	Sichuan	Shandong	Shanghai
abbreviation	SN	SX	TJ	XJ	XZ	YN
full name	Shaanxi	Shanxi	Tianjin	Xinjiang	Tibet	Yunnan
abbreviation	ZJ	AT	BE	BG	CY	CZ
full name	Zhejiang	Austria	Belgium	Bulgaria	Cyprus	Czech
abbreviation	DE	DK	EE	ES	FI	FR
full name	Germany	Denmark	Estonia	Spain	Finland	France
abbreviation	GR	HR	HU	IE	IS	IT
full name	Greece	Croatia	Hungary	Ireland	Iceland	Italy
abbreviation	LT	LU	LV	MT	NL	NO
full name	Lithuania	Luxembourg	Latvia	Malta	the Netherlands	Norway
abbreviation	PL	PT	RO	SE	SI	SK
full name	Poland	Portugal	Romania	Sweden	Slovenia	Slovakia

Table 1. Province/country abbreviation and full name.

(1) Combining with Life Cycle Assessment (LCA) methodologies to enable a comprehensive analysis of the full life cycle of EVs and their batteries, from raw material extraction and manufacturing to use and end-of-life treatment. This facilitates accurate carbon footprint calculations and assessments of the decarbonization potential of further EV deployment across sub-regions.
(2) Combining with the development of battery technologies and circular economy to estimate sub-regional demand and recycling potential for critical materials (such as lithium, cobalt, and nickel), and to support dynamic tracking of supply–demand trends in critical materials, offer insights into the stability and sustainability of battery supply chains and provide data support for relevant policy and decision-making.
(3) Analyzing the potential relationships between sub-regional EV market characteristics and their sustainability indicators (e.g., decarbonization performance and critical material demand), as well as the corresponding policies, to identify leading sub-regions in the electrification progress, extract lessons from their policy patterns, and assess the feasibility of scaling those strategies to other sub-regions.
(4) Extending the system boundary of the bottom-up, multi-dataset fusion approach to countries or regions beyond China and the EU, to enable them to track EV characteristics (including but not limited to those covered in this study), so as to better understand their sustainability performance and improvement opportunities.

Methods

This study uses a bottom-up method to analyze EV market characteristics, as shown in Fig. 2. First, we collect data on trim-level sales and specifications (trim denotes a specific version of a vehicle model with a distinct set of specifications, such as Tesla Model 3 Performance) for all Battery Electric Vehicle (BEV) and Plug-in Hybrid Electric Vehicle (PHEV) models available in both Chinese and EU markets in 2023 from several data sources. After processing and completing the missing information (e.g., battery capacity for certain trims), we organize the data into $60 \times 5 \times 12$ meta-categories, which reflect the combinations of (1) spatial (31 provinces in mainland China excluding Hong Kong, Macao, and Taiwan; 29 European countries covering EU27, Norway, and Iceland, with abbreviations and full names for each sub-region provided in Table 1), (2) segmental (5 segments categorized based on wheelbase), and (3) powertrain (BEV/PHEV with 6 battery chemistries) classifications. Based on this, we calculate market characteristics including EV market penetration, segment mix, battery chemistry mix, and sales-weighted average battery capacity for each sub-region and for the overall markets.

System boundary. This study covers electric passenger vehicles registered in 2023 across 31 provinces, municipalities, and autonomous regions in mainland China (excluding Hong Kong, Macao, and Taiwan) and 29 European countries (EU27 plus Norway and Iceland). "Passenger vehicles" here refer to vehicles used for passenger transport with no more than nine seats (including the driver seat), in line with the M1 classification used in both China and the EU[18,19]. This study focuses on BEVs and PHEVs, and does not include Fuel Cell Electric Vehicles (FCEVs), whose battery capacity in passenger vehicles is typically a negligible 2 kWh[12] (by comparison, the average battery capacities in the EU are 67 kWh for BEVs and 16 kWh for PHEVs). Combined with their limited global sales of only 8,000 units in 2023[20], the resource and environmental impacts of passenger FCEVs remain negligible.

For Hybrid Electric Vehicles (HEVs), the Nickel-Metal Hydride (NiMH) batteries they predominantly use[21] are similarly small in capacity. Although HEVs typically host a high sales share (accounting for a higher sales

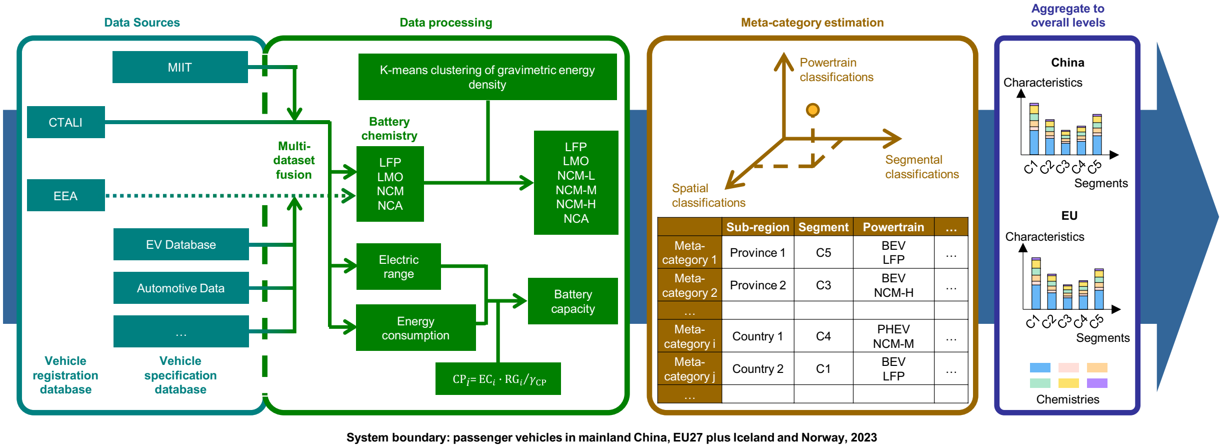

Fig. 2 Schematic diagram of the bottom-up approach to realize high-resolution EV market characterization based on trim data. Obtained from several databases, trim-level vehicle registration and specification data are processed and supplemented according to the equations and assumptions detailed in subsequent sections. Each trim entry is assigned to one of the 60 × 5 × 12 meta-categories according to its place of registration, vehicle segment, and powertrain. From the meta-category estimations, the EV market penetration, battery chemistry mix, and sales-weighted average battery capacity for each sub-region can be derived. On this basis, we get the overall EV market characteristics for China and the EU. The equation in the figure is detailed in Eq. (4) in the main text. MIIT, the Ministry of Industry and Information Technology; CTALI, China's Compulsory Traffic Accident Liability Insurance; EEA, European Environment Agency; EV, Electric Vehicle; BEV, Battery Electric Vehicle; PHEV, Plug-in Hybrid Electric Vehicle; LFP, Lithium iron phosphate battery; LMO, Lithium manganese oxide battery; NCM, Lithium nickel-cobalt-manganese oxide battery; NCM-L, Low-nickel content lithium nickel-cobalt-manganese oxide battery; NCM-M, Medium-nickel content lithium nickel-cobalt-manganese oxide battery; NCM-H, High-nickel content lithium nickel-cobalt-manganese oxide battery; NCA, Nickel-cobalt-aluminum oxide battery.

share than the combined total of BEVs and PHEVs in the EU in 2023[22]), their overall battery and nickel demand remains comparatively low. Moreover, HEVs remain primarily powered by fossil fuels and are often categorized as Internal Combustion Engine Vehicles (ICEVs)[12]. Therefore, this study also treats HEVs as ICEVs and does not characterize their batteries. Nonetheless, both HEVs and FCEVs remain important areas of research, and excluding them may limit the scope of insights into emerging technologies with evolving battery specifications. High-resolution data on these vehicles could better support comprehensive transport decarbonization studies.

Calculations. Each vehicle trim is assigned to a specific meta-category based on its registration location, vehicle segment, and the combination of powertrain and battery chemistry, corresponding to the spatial, segmental, and powertrain classifications defined in this study, as shown in Table 2. For any meta-category, total vehicle sales are calculated by summing up the sales of all vehicle trims within the category, as shown in Eq. (1). The market characteristics, such as battery capacity, are determined as the sales-weighted average of the specifications of the vehicle trims within the category, as depicted in Eq. (2). If needed, characteristics at broader spatial, segmental, and powertrain levels can be obtained by averaging across the appropriate categories, as shown in Eq. (3).

$$\mathrm{SA}_c = \sum_{i \in c} \mathrm{SA}_i \tag{1}$$

$$\mathrm{CHR}_c = \frac{\sum_{i \in c} \mathrm{SA}_i \cdot \mathrm{CHR}_i}{\sum_{i \in c} \mathrm{SA}_i} \tag{2}$$

$$\mathrm{CHR}_C = \frac{\sum_{c \in C} \mathrm{SA}_c \cdot \mathrm{CHR}_c}{\sum_{c \in C} \mathrm{SA}_c} \tag{3}$$

where,

SA_i is the sales of vehicle trim i;
CHR_i is any attribute or specification associated with vehicle trim i;
c is a vehicle meta-category, classified by region, segment, and powertrain as defined in this study;
C is the vehicle category at any broader aggregation level.

Data sources. To support these calculations, we use multiple data sources, as shown in Fig. 3.

Dimension	Classification	Code/Abbreviation	Notes
Spatial—China	31 Chinese provinces and municipalities	Abbreviations using province initials	Referring to the region where the vehicle is registered. The correspondence between abbreviations and full names is provided in Table 1.
Spatial—the EU	27 EU member states + Norway, Iceland	alpha-2 country code specified in ISO 3166	
Vehicle segment*	Mini	C1	Wheelbase ≤ 2450 mm
	Small	C2	2450 mm < Wheelbase ≤ 2600 mm
	Medium	C3	2600 mm < Wheelbase ≤ 2750 mm
	Large	C4	2750 mm < Wheelbase ≤ 2900 mm
	Executive	C5	2900 mm < Wheelbase
Powertrain—overall configuration	Internal combustion engine vehicle	ICEV	
	Plug-in hybrid electric vehicle	PHEV	Including extended-range electric vehicles
	Battery electric vehicle	BEV	
	Fuel cell electric vehicle	FCEV	Not considered in this study
Powertrain—battery chemistry**	Lithium iron phosphate battery	LFP	Battery chemistry is distinguished by the cathode material. The anode material is graphite for all categories
	Lithium manganese oxide battery	LMO	
	Low-nickel content lithium nickel-cobalt-manganese oxide battery (N:C:M = 1:1:1)	NCM-L	
	Medium-nickel content lithium nickel-cobalt-manganese oxide battery (N:C:M = 6:2:2 or 5:2:3)	NCM-M	
	High-nickel content lithium nickel-cobalt-manganese oxide battery (N:C:M = 8:1:1 or 9:0.5:0.5)	NCM-H	
	Nickel-cobalt-aluminum oxide battery	NCA	

Table 2. Classification standards for passenger vehicles. *This study defined vehicle segments by integrating common classification practices in China and Europe. In China, segments are typically based on wheelbase or length thresholds[33], while in Europe, they are often determined by comparison with benchmark models, without clear definitions[34,35]. Using collected wheelbase data, we tested evenly spaced thresholds and adjusted them iteratively until the models in each segment aligned reasonably with both Chinese and European conventions, thereby determining the final wheelbase thresholds. **Following common practice, Lithium nickel-cobalt-manganese oxide (NCM) batteries are further grouped based on their N:C:M ratios into two[5] or three[36,37] categories. This study applies the three-group scheme.

(1) **China's Compulsory Traffic Accident Liability Insurance (CTALI)**[23,24]: This database provides China's passenger vehicle registration data at the prefecture-city level, amounting to around 25 million entries annually, which we aggregate to the provincial level for analysis.
(2) **Ministry of Industry and Information Technology of the People's Republic of China (MIIT):** This database provides specifications for over 20,000 EV trims in total, including battery capacity, battery chemistry, gravimetric energy density, energy consumption rate, etc. Specifically, trim specifications were obtained from the MIIT's Policy Document Library[25] by searching the four types of catalogues published in various batches across different years: (a) Catalogue of New Energy Vehicles Exempted from Vehicle Purchasing Tax; (b) Catalogue of New Energy Vehicle Models with Admission to Vehicle and Vessel Tax Reduction and Exemption; (c) Catalogue of Recommended New Energy Vehicle Models for Promotion and Application; (d) Key Specifications of Recommended New Energy Vehicle Models for Promotion and Application.
(3) **European Environment Agency (EEA)**[26]**:** Containing data on nearly 10 million vehicle entries annually, it provides a detailed description of passenger vehicles registered in the EU, Norway, and Iceland, including country of registration, manufacturer name, commercial name of the vehicle, vehicle variant, vehicle version, wheelbase, fuel mode, and sales. The data presented in this paper is based on the final version of EEA's 2023 dataset, released in December 2024.
(4) **Online EV specification databases:** including but not limited to EV Database (https://ev-database.org) and Automotive Data (https://www.auto-data.net/en). They are used to help supplement missing specifications in other aforementioned databases, such as the EEA.

Data processing. Because these datasets are independent and often incomplete, we apply several steps to clean, supplement, and combine them, as shown in Fig. 4.

(1) **Estimating missing battery capacities (MIIT)**
For a few trims lacking battery capacity values, we estimate them using energy consumption rate and electric range, as shown in Eqs. (4, 5).

China

CTALI

		Registrations											
	VAC	Provincial level	AH			BJ	…	YN			ZJ		
		Prefecture-city level	HFE	WHI	…	BJ	…	KMG	QJS	…	HGH	NGB	WNZ
Trim 1	●		●	●		●		●	●		●	●	●
Trim 2	●		●	●		●		●	●		●	●	●
Trim 3	●		●	●		●		●	●		●	●	●
…													
Trim i	●		●	●		●		●	●		●	●	●
Trim j	●		●	●		●		●	●		●	●	●
…													

MIIT

	VAC	Battery capacity (kWh)	Powertrain	Electric range (km)	Energy consumption (Wh/km)	Gravimetric energy density (Wh/kg)	…
Trim 1	●	●	BEV-LFP	●	●	●	
Trim 2	●	●	PHEV-NCM	●	●	●	
Trim 3	●	●	BEV-NCM	●	●	●	
…							
Trim i	●	○	BEV-LMO	●	●	●	
Trim j	●	○	BEV-NCA	●	●	●	
…							

EU

EEA

	Model	Variant	Electric range (km)	Energy consumption (Wh/km)	Registrations					
					Country level	AT	BE	…	SI	SK
Trim 1	●	●	●	●		●	●		●	●
Trim 2	●	●	●	○		●	●		●	●
Trim 3	●	●	○	●		●	●		●	●
…										
Trim i	●	●	○	○		●	●		●	●
Trim j	●	●	○	○		●	●		●	●
…										

Online EV specification databases: EV Database, Automotive Data…

	Model	Variant	Battery capacity (kWh)	Powertrain	Electric range (km)	Energy consumption (Wh/km)	Gravimetric energy density (Wh/kg)	…
Trim 1	●	●	●	BEV-NCM811	●	●	○	
Trim 2	●	○	●	PHEV-NCM	●	●	●	
Trim 3	●	●	●	BEV-NCA	●	●	○	
…								
Trim i	●	○	●	BEV-NCM	●	●	●	
Trim j	●	○	●	BEV-NCM523	●	●	○	
…								

○ Unknown ● Known

Fig. 3 Illustration of data source content. Both China and the EU rely on two datasets respectively. The upper dataset mainly records vehicle registration data: CTALI for China and the EEA database for the EU. The lower dataset provides vehicle specification data: MIIT for China and several online sources such as EV Database for the EU. Hollow dots indicate missing fields for a given entry, while solid dots indicate that the corresponding fields are available. In China, each trim's VAC is known in both datasets, allowing for direct integration via VAC matching. In the EU, trims largely correspond to "variants" in the EEA database, but variant information is only available for a subset of trims in the specification databases. Therefore, when variant information is available in both datasets, matching is performed directly at the variant level. If the variant is only available in the EEA database but missing in the specification databases, we first match at the model level (using the model's name to locate candidate trims in the specification databases), then use attributes such as electric range and energy consumption from the EEA database to identify the most likely corresponding trim, as detailed in the Data processing section. VAC, Vehicle Announcement Code; BEV, BEV, Battery Electric Vehicle; PHEV, Plug-in Hybrid Electric Vehicle; LFP, Lithium iron phosphate battery; LMO, Lithium manganese oxide battery; NCM, Lithium nickel-cobalt-manganese oxide battery; NCA, Nickel-cobalt-aluminum oxide battery; AH, Anhui; BJ, Beijing, YN, Yunnan; ZJ, Zhejiang; HFE, Hefei; WHI, Wuhu; KMG, Kunming; QJS, Qujing; HGH, Hangzhou; NGB, Ningbo; WNZ, Wenzhou; AT, Austria; BE, Belgium; SI; Slovenia; SK, Slovakia.

$$\mathrm{CP}_i = \frac{\mathrm{EC}_i \cdot \mathrm{RG}_i}{\gamma_{\mathrm{CP}}} \tag{4}$$

$$\gamma_{\mathrm{CP}} = \frac{\sum_{j \in J} \frac{\mathrm{EC}_j \cdot \mathrm{RG}_j}{\mathrm{CP}_j}}{n_J} \tag{5}$$

where,

i is the trim without battery capacity information;
j is the trim with battery capacity information;
EC_i and EC_j are the energy consumption rates of trim i and trim j (kWh/km);
RG_i and RG_j are the electric range of trim i and trim j (km);
CP_j is the battery capacity of trim j (kWh);
n_J is the number of all trims with battery capacity information;
γ_{CP} is the percentage of battery energy that can be actually used out of the total battery capacity (%), which is calculated separately for BEVs and PHEVs.

(2) **Distinguishing sub-NCM chemistries (MIIT)**
The battery chemistry provided in the MIIT database is only detailed to the NCM level, with no distinctions among NCM-L, NCM-M, and NCM-H batteries. To address this, we classify NCM batteries into three groups based on their pack-level gravimetric energy density (provided by the MIIT database) using the K-means approach[27], with the lowest to highest gravimetric energy density groups aligned to NCM-L,

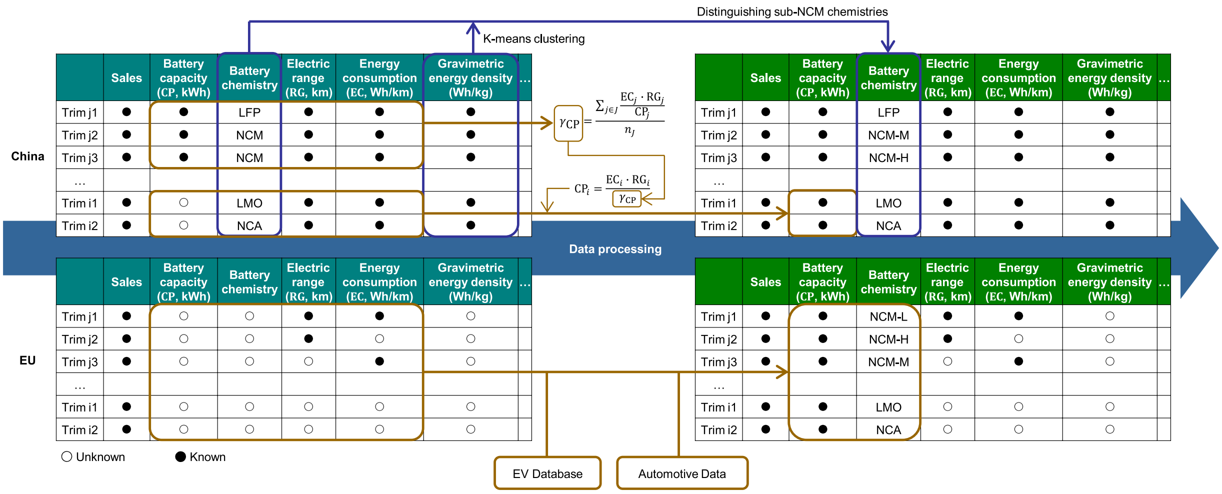

Fig. 4 Illustration of data processing. On the China side, data processing involves two main tasks: supplementing missing battery capacities and determining sub-NCM chemistries based on energy density. For the first task, we first establish the relationship between battery capacity, electric range, and energy consumption using entries where all three fields are available. This relationship is then applied to entries with missing battery capacity but known electric range and energy consumption, as illustrated by the two equations, which are actually Eqs. (4, 5) in the main text. For the second task, we perform K-means clustering on pack-level gravimetric energy densities. Since higher nickel content corresponds to higher energy density, the resulting clusters are ordered by energy density and classified into NCM-L, NCM-M, and NCM-H accordingly. The clustering process is carried out separately for BEVs and PHEVs. On the EU side, we use available information in the EEA dataset such as model name, electric range, and energy consumption, to identify the most likely corresponding trim in specification sources such as EV Database and Automotive Data. Once the trim is identified, we then supplement the corresponding entry with battery capacity and battery chemistry information drawn from these external sources. Hollow dots indicate missing fields for a given entry, while solid dots indicate that the corresponding fields are available. LFP, Lithium iron phosphate battery; LMO, Lithium manganese oxide battery; NCM, Lithium nickel-cobalt-manganese oxide battery; NCM-L, Low-nickel content lithium nickel-cobalt-manganese oxide battery; NCM-M, Medium-nickel content lithium nickel-cobalt-manganese oxide battery; NCM-H, High-nickel content lithium nickel-cobalt-manganese oxide battery; NCA, Nickel-cobalt-aluminum oxide battery.

NCM-M, and NCM-H categories, respectively. This process is carried out separately for BEVs and PHEVs because of their different energy density attributes. Taking the clustering of BEVs as an example, the energy density data used for clustering (BEV_NCM_energy_density_CN.xlsx), the corresponding MATLAB clustering code (BEV_Energy_density_clustering_CN.m), and the clustering results (BEV_NCM_clustered_CN.xlsx) are all available on Figshare[28]. The clustering results are shown in Fig. 5.

It is worth noting that identifying sub-NCM chemistries based on energy density clustering rather than actual elemental composition may occasionally lead to mismatches. For example, certain EVs equipped with NCM523 cells (which should be classified as NCM-M) might be categorized as NCM-L if their battery packs have unusually low energy density due to inefficient packing. Nevertheless, given the large volume of trim data in China and the lack of reliable databases distinguishing sub-NCM chemistries, clustering based on energy density remains a relatively robust and efficient approach, and has also been adopted in previous studies[29].

(3) **Supplementing battery capacity and battery chemistry (EEA)**
The EEA database does not directly report battery capacity or battery chemistry, so we supplement both using external data sources such as EV Database and Automotive Data, as well as the technical data documents from the Original Equipment Manufacturers (OEMs). We match the vehicle trims with external battery capacity and chemistry data by following the rules below.

(a) In the EEA database, each entry is defined with three levels of vehicle identification, i.e., model (represented by its commercial name), variant, and version. In most cases, the “variant” aligns with the “trim” definition in this study, though in some cases “model” or “version” does instead. Due to the challenge of decoding the variant or version information for many entries, vehicle characteristics are matched to the “model” level in most circumstances. Most Tesla models, as well as certain models from other OEMs such as BMW i3, are exceptions. Their entries can be decoded to the variant level by referring to publicly available resources. For most manufacturers, while typically there are numerous variants under one model, the battery characteristics are largely consistent for the variants of the same model. Therefore, decoding only to the model level does not introduce significant bias in most cases.

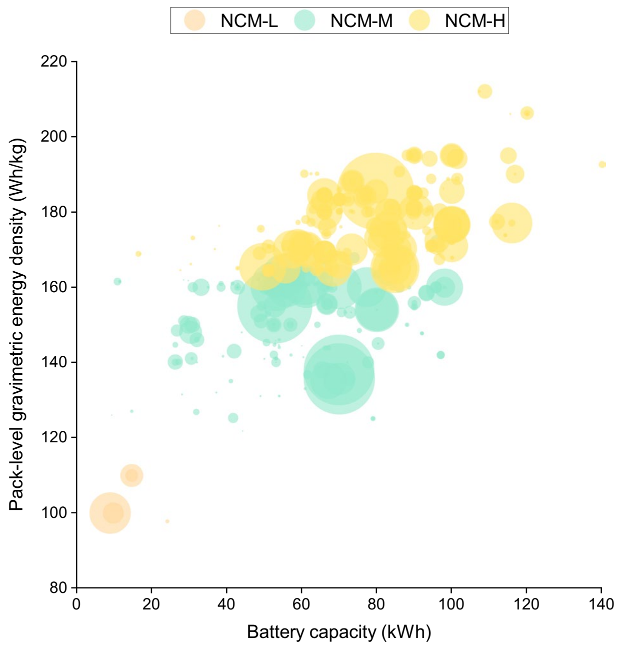

Fig. 5 Visualization of BEV clustering results in China. Vehicles equipped with NCM batteries are grouped into three categories based on the pack-level gravimetric energy density of their batteries using the K-means clustering approach, with each category represented in a different color. Each dot represents a VAC from the MIIT database, which aligns with the "trim" defined in this study. The size of each dot corresponds to the total nationwide sales of the trim in 2023.

(b) For entries in the EEA database with available energy consumption rate and/or electric range, we compare these values with those from external sources to identify the most likely battery capacity when multiple options exist.
(c) For entries lacking both energy consumption rate and electric range, where the associated model includes multiple trims with different battery capacities but the specific trim of the entry cannot be identified, the intermediate (for more than 3 capacities) or largest (for 2 capacities) battery capacity value is used.
(d) For entries whose battery chemistries can be externally identified as NCM, but whose sub-NCM chemistries remain unknown, we refer to the clustering results from the dataset of EVs in China to assign sub-NCM chemistries, provided that their gravimetric energy densities are available externally.
(e) If battery chemistry or capacity of certain trims remains unavailable after complete searches from various sources, such trims are not included in the statistics. In this study, 6.8% of entries fall into this category. The trims in these entries, with generally very low sales, accounted for only 1.6% of total sales, meaning that the neglect of these trims has a marginal impact on the sales-weighted characteristics.

Data Records

The dataset is available for download from Figshare[28], comprising two workbooks (CN_characteristics_2023.xlsx and EU_characteristics_2023.xlsx), which compile the EV market characteristics in China, the EU, and their respective sub-regions presented in this study. The contents of each worksheet and the meaning of its columns are detailed in Table 3.

Using the EU as an example, we also provide code for calculating EV market characteristics based on the bottom-up, multi-dataset fusion approach (EU_Bottom_up_Multi_dataset_fusion.m), written in MATLAB R2021b. Given two inputs: (1) the 2023 EU passenger vehicle registration data in comma-separated values (.csv) format (referred to as "EEA_Y2023.csv" in the code; not provided on Figshare due to file size, but directly available for download from the EEA database[26]), and (2) a complete EU EV specification table organized at the "variant" level based on the EEA database, the code can output the EV characteristics presented in this study. A sample EU EV specification file (EU_Variant_2023.xlsx) is also provided on Figshare, covering EVs that together account for nearly 70% of total EU sales in 2023. This file includes input data used for multi-dataset fusion, such as vehicle segment, energy consumption, electric range, battery capacity, and battery chemistry, as well as 2023 registrations of each variant at both the EU and individual country levels, which are the direct output of the fusion process.

A counterpart file of EU_Variant_2023.xlsx for China, CN_VAC_2023.xlsx, is organized at the Vehicle Announcement Code (VAC) level based on the MIIT database and is also available on Figshare. The registration data in this sample file also reflect the direct outputs of the fusion process, which for China was conducted at the prefecture-level city resolution. To prevent the table from becoming overly lengthy and complex, we present only the vehicle registrations of each provincial capital as a representative subset. Besides the same set of information as in EU_Variant_2023.xlsx, it also provides battery energy density and vehicle curb weight. As this

Sheet	Column	Description
EV_characteristics ICEV_and_EV_Inflows	Sub-region	The provinces/municipalities in China or the member states in the EU. Sub-regions are sorted in ascending order and represented by their two-letter abbreviations, with the full names provided in Table 1. Aggregate values for China and the EU are also included and denoted as “CN” and “EU”, respectively.
	Segment	Vehicle segment, as defined in Table 1. Aggregated results across all segments are also included and denoted as “ALL”.
	Powertrain	Twelve powertrains derived from the combination of the two overall configurations with the six battery chemistries (e.g., BEV-LFP, BEV-NCA). Aggregated results across all battery chemistries are also included and denoted as “BEV-ALL” and “PHEV-ALL”.
EV_characteristics	Inflow	New EV registrations specified by sub-region, segment, and powertrain.
	Segment_mix	The proportion of new EV registrations in a specified segment relative to the total new EV registrations across all segments within a given sub-region and powertrain. The value can be calculated based on the Inflow values across all segments. For example, if the Segment_mix value for Segment “C1” under the Powertrain “BEV-LFP” in Sub-region “BJ” is 0.6, it means that 60% of newly registered BEVs powered by LFP batteries in Beijing were in the mini (C1) segment. By definition, when the Segment is “ALL”, the Segment_mix value is always 1.
	Chemistry_mix	The proportion of new EV registrations with a specified battery chemistry relative to the total new EV registrations across all battery chemistries within a given sub-region, segment, and overall powertrain configuration (BEV or PHEV). The value can be calculated based on the Inflow values across all powertrains under the same overall configuration (BEV or PHEV). For example, if the Chemistry_mix value for the Powertrain “BEV-LFP” in Sub-region “BJ” and Segment “C1” is 0.6, it indicates that 60% of newly registered BEVs in the mini (C1) segment in Beijing are powered by LFP batteries. By definition, when the Powertrain is “BEV-ALL” or “PHEV-ALL”, the Chemistry_mix value is always 1.
	Capacity_kWh	Sales-weighted average battery capacity for a given sub-region, segment, and powertrain, measured in kilowatt-hours (kWh). It is calculated based on the battery capacities of all newly registered vehicles, weighted by their respective registration volumes. A value of 0 indicates that no such vehicles were registered.
ICEV_and_EV_Inflows	Total_inflows	New passenger vehicle registrations specified by sub-region and segment.
	ICEV_inflows	New ICEV registrations specified by sub-region and segment.
	BEV_inflows	New BEV registrations specified by sub-region and segment.
	PHEV_inflows	New PHEV registrations specified by sub-region and segment.
	BEV_penetration	Percentage of new BEV registrations among new passenger vehicle registrations in a given sub-region and segment.
	PHEV_penetration	Percentage of new PHEV registrations among new passenger vehicle registrations in a given sub-region and segment.

Table 3. Detailed description of the dataset.

study focuses on three key characteristics: vehicle segment, battery capacity, and battery chemistry, which are most important for EV-related environmental analyses such as critical material demand and decarbonization potential assessments, we strived to ensure complete data for these characteristics in every entry, including by supplementing from external sources when necessary. For other fields such as energy consumption and electric range, we only recorded information directly provided by the basic source databases (MIIT or EEA) and did not attempt to complete missing data through additional searches.

Technical Validation

The results from this bottom-up method are validated by comparing them to existing top-down estimations. Specifically, country-level EV market penetration rates derived in this study are compared with those reported in the International Energy Agency’s Global EV Data Explorer[6], as shown in Table 4. China and the eight European countries with the largest EV markets (ranked by EV market penetration rate in descending order) were selected for comparative validation of EV penetration in 2023. Compared with results from IEA’s Global EV Data Explorer[6], most countries have a relative deviation of no more than 5%. The deviation is mainly rooted in the differences in statistical scopes. The results from the IEA’s Global EV Data Explorer are specifically for the “Cars” category among its four classifications: Buses, Cars, Trucks, and Vans. The scope of our study is electric passenger vehicles (EPV), including sedans, SUVs, MPVs, and Minivans. The two scopes mostly overlap, but also have different elements. For example, the minivan segment is included in our EPV scope in our study, but not in the IEA’s Cars scope.

The battery installation capacity of different battery chemistries for electric passenger vehicles in China are compared with existing top-down estimates, as shown in Table 5. Due to the absence of similar top-down battery installation capacities for the EU that align with the statistical scope of this paper, the validation is limited to China and conducted using data from the China Automotive Battery Innovation Alliance (CABIA)[30], which reports battery installation related to EV production in China, categorized by battery chemistry (LFP vs. NCM/NCA), and by vehicle type (passenger vehicles vs. commercial vehicles). We trim the CABIA estimations to be comparable with our estimation in the following steps. (1) Given that commercial vehicles in China rarely use NCM or NCA batteries and predominantly rely on LFP batteries (accounting for 90% of batteries in electric trucks and buses[20], and 99.9% in heavy-duty vehicles in 2023[31]), we assume that the NCM and NCA battery installations are entirely attributed to passenger vehicles, which may introduce limited uncertainty due to the already dominant shares noted above. Since other battery chemistries account for less than 1% of total installations, we approximate LFP battery installations in passenger vehicles as the total passenger vehicle battery

Country	This study	IEA	Relative deviation
China	34.2%	38.0%	−10%
Norway	90.4%	90.0%	0.4%
Sweden	59.9%	60.0%	−0.2%
The Netherlands	43.9%	44.0%	−0.2%
Belgium	40.7%	41.0%	−0.7%
France	25.1%	25.0%	0.4%
Germany	25.3%	24.0%	5.4%
Spain	12.1%	12.0%	0.8%
Italy	8.7%	9.2%	−5.4%

Table 4. Comparison of EV penetration between this study and the IEA.

Chemistry	This study	CABIA	Relative deviation
All	328	305	7.5%
LFP	203	196	3.6%
NCM and NCA	125	109	14.7%

Table 5. Comparison of battery installation in China between this study and CABIA.

installations minus NCM and NCA battery installations. (2) We convert the production-based battery installation to sales-based battery installation by using the ratio of domestic EPV sales to production volume from the China Automobile Dealers Association (CADA)[32]. However, since the ratio of sales to production volume (quantity-to-quantity) provided by CADA does not precisely reflect the ratio of production-based installation capacities to sales-based installation capacities (capacity-to-capacity), using these top-down estimates for validation would still introduce discrepancies. Furthermore, although not explicitly stated by CABIA, its definition of passenger vehicles may differ from that adopted in this study. Despite these differences, in most cases, the relative deviation between the results obtained using the two approaches remains within 15%, which is considered acceptable.

Code availability

The clustering code and the bottom-up multi-dataset fusion code for characteristic calculation (using EU-side data as an example) were written using MATLAB R2021b. The visual plots were realized through Origin 2025. The purpose of each script is mentioned in the comments, and the code was uploaded to Figshare along with the dataset[28].

Received: 25 June 2025; Accepted: 4 August 2025;
Published online: 19 August 2025

References

1. Zhang, H., Hu, X., Hu, Z. & Moura, S. J. Sustainable plug-in electric vehicle integration into power systems. *Nature Reviews Electrical Engineering* **1**, 35–52 (2024).
2. Deng, J., Bae, C., Denlinger, A. & Miller, T. Electric Vehicles Batteries: Requirements and Challenges. *Joule* **4**, 511–515 (2020).
3. Liang, X. *et al.* Air quality and health benefits from fleet electrification in China. *Nature Sustainability* **2**, 962–971 (2019).
4. Muratori, M. *et al.* The rise of electric vehicles—2020 status and future expectations. *Progress in Energy* **3** (2021).
5. *Global EV Outlook 2025*; https://www.iea.org/reports/global-ev-outlook-2025 (International Energy Agency, 2025).
6. International Energy Agency. Global EV Data Explorer. https://www.iea.org/data-and-statistics/data-tools/global-ev-data-explorer (2025).
7. Marianne Kah, S. L., Chiu, J., Wong, H. X. *Forecasts of Electric Vehicle Penetration and its Impact on Global Oil Demand*; https://www.energypolicy.columbia.edu/publications/forecasts-electric-vehicle-penetration-and-its-impact-global-oil-demand/ (2022).
8. Deng, Y., Hao, H. & Jia, C. Exploring the potential of cutting battery use in electric vehicles. *Clean Technologies and Environmental Policy* **26**, 367–379 (2023).
9. Vega-Perkins, J., Newell, J. P. & Keoleian, G. Mapping electric vehicle impacts: greenhouse gas emissions, fuel costs, and energy justice in the United States. *Environmental Research Letters* **18**. https://doi.org/10.1088/1748-9326/aca4e6 (2023).
10. Llamas-Orozco, J. A. *et al.* Estimating the environmental impacts of global lithium-ion battery supply chain: A temporal, geographical, and technological perspective. *PNAS Nexus* **2**. https://doi.org/10.1093/pnasnexus/pgad361 (2023).
11. Cheng, A. L., Fuchs, E. R. H., Karplus, V. J. & Michalek, J. J. Electric vehicle battery chemistry affects supply chain disruption vulnerabilities. *Nat Commun* **15**, 2143 https://www.ncbi.nlm.nih.gov/pubmed/38459029 (2024).
12. Zhang, C., Zhao, X., Sacchi, R. & You, F. Trade-off between critical metal requirement and transportation decarbonization in automotive electrification. *Nat Commun* **14**, 1616, https://doi.org/10.1038/s41467-023-37373-4 (2023).
13. Global EV Sales for 2023. *EV-Volumes* https://ev-volumes.com/news/ev/global-ev-sales-for-2023/ (2024).
14. *Global EV Outlook 2023*; https://www.iea.org/reports/global-ev-outlook-2023 (International Energy Agency, 2023).
15. Ren, Y. *et al.* Hidden delays of climate mitigation benefits in the race for electric vehicle deployment. *Nature Communications* **14**, 3164, https://doi.org/10.1038/s41467-023-38182-5 (2023).
16. Lu, P., Hamori, S., Sun, L. & Tian, S. Does the electric vehicle industry help achieve sustainable development goals?—evidence from China. *Frontiers in Environmental Science* **11** (2024).

17. Hao, X., Zhou, Y., Wang, H. & Ouyang, M. Plug-in electric vehicles in China and the USA: a technology and market comparison. *Mitigation and Adaptation Strategies for Global Change* **25**, 329–353, https://doi.org/10.1007/s11027-019-09907-z (2020).
18. Standardization Administration of the People's Republic of China. GB/T 15089: Classification of power-driven vehicles and trailers (2001).
19. EU classification of vehicle types. *European Alternative Fuels Observatory* https://alternative-fuels-observatory.ec.europa.eu/general-information/vehicle-types (2024).
20. *Global EV Outlook 2024*; https://www.iea.org/reports/global-ev-outlook-2024 (International Energy Agency, 2024).
21. Edmondson, J. Hybrid Electric Vehicles: A Stay of Execution for NiMH Batteries. *IDTechEX* https://www.idtechex.com/tw/research-article/hybrid-electric-vehicles-a-stay-of-execution-for-nimh-batteries/22786 (2021).
22. *New car registrations: +13.9% in 2023; battery electric 14.6% market share*; https://www.acea.auto/files/Press_release_car_registrations_full_year_2023.pdf (the European Automobile Manufacturers' Association, 2024).
23. Automotive Data of China Co., Ltd. *China Automotive Low Carbon Action Plan (2022): Low Carbon Development Strategy and Transformation Path for Carbon Neutral Automotive* (China Machine Press, Beijing, 2023).
24. Auto Data Center of CATARC, accessed 1 Aug 2022; http://www.catarc.info/.
25. *Ministry of Industry and Information Technology Policy Document Library* https://www.miit.gov.cn/search/wjfb.html (2024).
26. Monitoring of CO2 emissions from passenger cars. *European Environment Agency* https://www.eea.europa.eu/en/datahub/datahubitem-view/fa8b1229-3db6-495d-b18e-9c9b3267c02b (2024).
27. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences* **622**, 178–210 (2023).
28. Mai, L., Dong, Z., Li, H., Geng, J. & Deng, Y. A high-resolution dataset on the electric passenger vehicle characteristics in China and the European Union. *Figshare* https://doi.org/10.6084/m9.figshare.29683073 (2025).
29. Baars, J., Domenech, T., Bleischwitz, R., Melin, H. E. & Heidrich, O. Circular economy strategies for electric vehicle batteries reduce reliance on raw materials. *Nature Sustainability* **4**, 71–79, https://doi.org/10.1038/s41893-020-00607-0 (2020).
30. Monthly information. *China Automotive Battery Innovation Alliance* https://mp.weixin.qq.com/mp/appmsgalbum?__biz=MzU3ODQ4MDg5Mw==&action=getalbum&album_id=3286101822386192388#wechat_redirect (2024).
31. Jin, L. & Mao, S. *Zero-emission bus and truck market in China in 2023*; https://theicct.org/wp-content/uploads/2024/08/ID-191-%E2%80%93-EU-R2Z-Q1_final-1.pdf (the International Council on Clean Transportation, 2024).
32. *Statistics data*; http://data.cada.cn/main/overview.do (China Automobile Dealers Association, 2024).
33. Yang, J. How to distinguish the car level? *ruiyaone.com* https://ruiyaone.com/blogs/car-knowledge-lfotpp/how-to-distinguish-the-car-level (2019).
34. Regulation (EEC) No 4064/89 - Merger Procedure; https://ec.europa.eu/competition/mergers/cases/decisions/m1406_en.pdf (Office for Official Publications of the European Communities L-2985 Luxembourg, 1999).
35. Thiel, C., Schmidt, J., Van Zyl, A. & Schmid, E. Cost and well-to-wheel implications of the vehicle fleet CO2 emission regulation in the European Union. *Transportation Research Part A: Policy and Practice* **63**, 25–42, https://doi.org/10.1016/j.tra.2014.02.018 (2014).
36. Dong, Z. *et al*. Projecting future critical material demand and recycling from China's electric passenger vehicles considering vehicle segment heterogeneity. *Resources, Conservation and Recycling* **207**, 107691, https://doi.org/10.1016/j.resconrec.2024.107691 (2024).
37. Tran, Y. H. T., An, K., Vu, D. T. T. & Song, S.-W. High-Voltage Electrolyte and Interface Design for Mid-Nickel High-Energy Li-Ion Batteries. *ACS Energy Letters* **10**, 356–370, https://doi.org/10.1021/acsenergylett.4c02860 (2025).

Acknowledgements

This study is sponsored by the National Natural Science Foundation of China (72122010), National Key R&D Program of China (2019YFC1908501), supported by Carbon Neutrality and Energy System Transformation (CNEST) Program led by Tsinghua University, and International Joint Mission on Climate Change and Carbon Neutrality.

Author contributions

L.M. and M.L. conducted the analysis and wrote the paper. X.S., F.M., Z.W. and J.E.T. reviewed and edited the paper. Y.G. and H.H. provided guidance on the paper. H.H. supervised the work and secured funding for the project. L.M., Z.D., and H.L. provided data for the analysis. J.G., H.D. and Y.D. helped with the methodology. F.B., Z.L. and F.Z. reviewed the paper and provided suggestions.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.H.