nature communications

Article https://doi.org/10.1038/s41467-025-57217-7

Unused housing in urban China and its carbon emission impact

Received: 29 May 2023

Accepted: 14 February 2025

Published online: 26 February 2025

Check for updates

Hefan Zheng, Rongjie Zhang ✉, Xinru Yin & Jing Wu ✉

The intensive utilization of residential space is crucial to the transition to a carbon-neutral residential sector, although it has received limited attention in the literature. We develop a methodology to estimate the volume of unused housing in urban China, defined as dwelling units built and sold for at least two years but never occupied. By early 2021, 17.4% of the housing stock built in China during the first two decades of this century remained unused. The construction and operation of unused housing produce 55.81 million tons of carbon dioxide annually at the national level, accounting for 6.9% of the Chinese residential sector's carbon emissions or 26.5% of the carbon emission reductions achieved by China's primary ongoing residential decarbonization efforts. Cutting down the volume of unused dwelling units can contribute significantly to China's decarbonization in 2021–2030.

One-third of worldwide carbon emissions are attributable to the construction and operation of residential buildings[1], making the residential sector a key component of global carbon mitigation. Decarbonizing the residential sector is particularly challenging in developing countries (i.e., those classified by the International Monetary Fund as "emerging market and developing economies," based on factors such as per capita income, economic structure, and level of global integration) such as China, where the housing stock continues to rise rapidly. Between 2001 and 2020, China built 11.47 billion square meters (m^2) of urban housing, accounting for about half of the world's new housing[2].

The majority of current efforts to decarbonize the residential sector, in both China and high-income countries (those with a Gross National Income per capita above $14,005 as classified by the World Bank in 2025), are centered on the "efficiency" perspective, which seeks to reduce the carbon emission intensity associated with the life cycle of residential buildings through measures such as construction material substitution during the materialization stage[3,4] and energy efficiency improvement during the operation stage[5,6]. Nevertheless, the "sufficiency" perspective, which focuses on the intensive usage of residential buildings, plays an even greater role in the pathway to establishing a carbon-neutral residential sector[7–9]. For instance, studies based on the United States and other high-income countries imply housing size is the primary determinant of residential carbon emissions in their countries[7,10]. In the context of China, where the volume of newly built housing construction has remained at an increasingly high level during the past two decades, the sufficiency perspective concentrates on the potential oversupply associated with the extensive new construction[11,12]. However, a comprehensive assessment of underoccupied housing and its impact on carbon emissions is still relatively rare.

Researchers and policymakers typically adopt the term "vacancy" for all underoccupied housing held by households but without residents at the time of investigation. Nevertheless, various types of housing vacancy have vastly different implications for decarbonization: some dwelling units are only temporarily vacant during normal housing turnovers in the market[13,14], some units are occasionally vacant but still serve households' specific housing demand as seasonal or second homes[15], and some units were previously fully inhabited but have become obsolete due to deteriorating physical or neighborhood conditions[16]. Whether eliminating the aforementioned types of vacancies is feasible without impairing housing market efficiency or resident well-being remains an open question. Therefore, this study focuses on an extreme type of underoccupied housing in urban China that is closest to resource waste—dwelling units that remain unoccupied for an extended period after completion. Specifically, we define a dwelling unit as "unused housing" if it (1) has been sold to a household, (2) has been completed for at least two years (so that the owner/renter has sufficient time to decorate the interior and move in), and, most importantly, (3) has never been occupied by the time of investigation.

Department of Construction Management and Hang Lung Center for Real Estate, Tsinghua University, Beijing 100084, China.
✉e-mail: zhangrongjie@tsinghua.edu.cn; ireswujing@tsinghua.edu.cn

https://doi.org/10.1038/s41467-025-57217-7

From the standpoint of sustainable development, the so-defined unused housing has never been functional and, thus, should and can be mitigated, if not avoided altogether, when decarbonizing the residential sector.

No current statistics on the volume or proportion of such unused housing in China or any other country are available; therefore, we develop a deep-learning-based methodology. Note our estimate only covers unused units that have been sold to and held by households. The volume of unsold and unoccupied units held by housing developers, or the so-called "developers' inventory," has been regularly measured and publicly released by the National Bureau of Statistics of China[17] and widely discussed in the literature[18]. As conceptually illustrated in Fig. 1 and described in detail in "Methods," the method consists of two core procedures. The first step is to develop a supervised deep learning algorithm to utilize the visual information of online-listed dwelling units and identify all unused units in the full-sample online-listed unit observations in the target cities, achieving the proportion of unused units among online-listed units (i.e., the listing-based unused rate, or LUR). The second step is to convert the LUR to the stock-based unused rate (SUR, defined as the proportion of unused units throughout the entire housing stock in the target city), which is our main interest. We have also adopted multiple methods to validate the accuracy of the classification results, which we describe in detail in Supplementary Note 1.

Our estimate suggests that in early 2021, 17.4% of the housing stock completed between 2001 and 2018 in urban China remained unused. The unused rate is particularly high in cities or housing sectors that are more likely to have experienced a substantial housing oversupply during the last two decades, especially in the majority of third-tier cities (as defined in Supplementary Table 1, where the city tier classification follows the definition provided by the National Bureau of Statistics of China). Moreover, we are able to achieve an estimate of the remarkable volume of carbon emissions associated with the construction and operation of these unused housing units, which has substantially counteracted China's ongoing efforts to decarbonize the residential sector. Taking the year 2020 as an example, the construction and operation of unused housing resulted in the emission of 55.81 million tons of carbon dioxide (CO_2) at the national level, which accounts for 6.9% of the total residential-sector carbon emissions and is equivalent to around 26.5% of the carbon emission reductions achieved by the Chinese government's four major carbon-mitigating measures in the residential sector. Naturally, utilizing existing unused housing and avoiding its further expansion should be designated as a top priority in establishing a carbon-neutral residential sector in China. Our scenario analysis suggests the total carbon emissions of the Chinese residential sector in 2021–2030 can be reduced by 9.1% if the current unused rate is cut by half by 2030.

Results

Volume of unused housing in major cities

We collected information on all dwelling units listed between October 2020 and August 2021 on a leading and anonymous online housing listing platform in mainland China, covering 56 major cities (Supplementary Table 1). In 2020, these 56 cities accounted for 45.0% of the urban population, 53.4% of gross domestic product, and 48.4% of the urban housing completions in China. To further illustrate the spatial distribution of the sample cities, we also present them on a map in Supplementary Fig. 1. In each city, we focus on housing communities completed by real estate developers between 2001 and 2018, considering that China's real estate industry only emerged at the beginning of this century.

Figure 2a depicts the baseline estimate of the city-level SURs during the sample period (we take the midpoint of the sample period, early 2021, to represent the sample period in the following discussions). We also verify the reliability of the estimates based on other housing market indicators (Supplementary Note 1).

Generally, the results indicate a substantial portion of new homes completed during the first two decades of this century in urban China had never been occupied by early 2021. Using the city-level aggregated housing completions between 2001 and 2018 as the weight, the weighted average SUR of these 56 major cities reached 17.4%; that is, for every six dwelling units completed and sold to households between

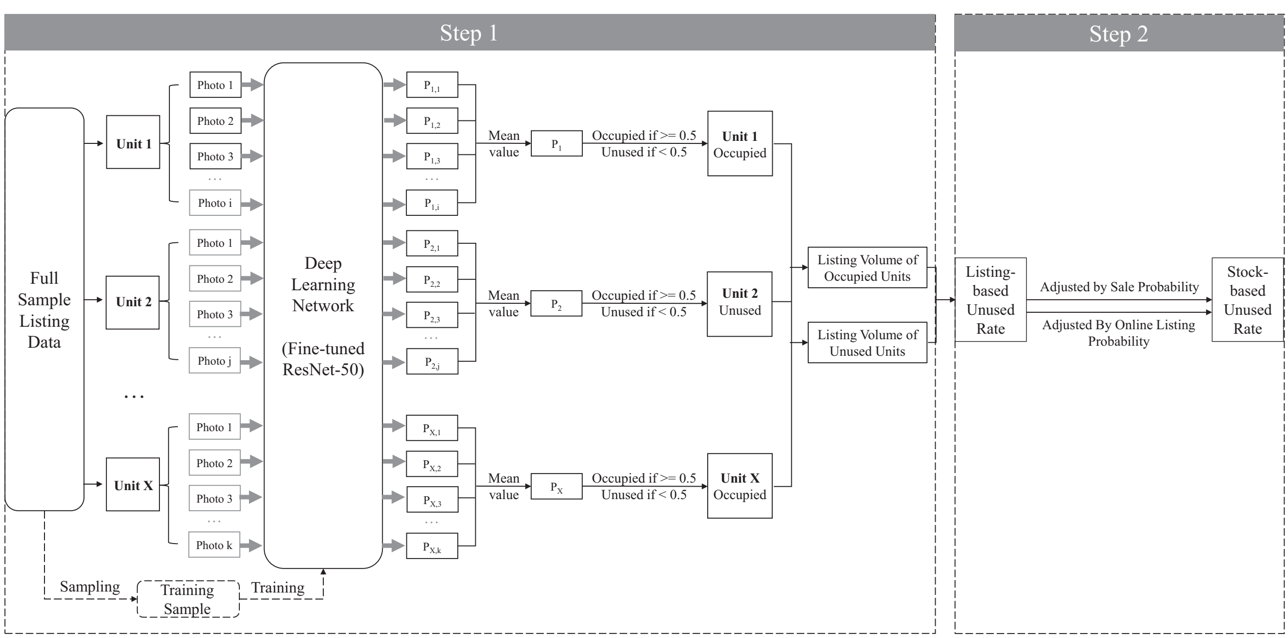

Fig. 1 | Method of estimating unused-housing rate. This figure displays our conceptual method of calculating the stock-based unused-housing rate based on the listing dataset. Step 1 is to present the method of identifying all the unused dwelling units for each sample city using a supervised deep learning network. Step 2 is to convert the listing-based unused rate into the stock-based unused rate. $P_{1,1}$ denotes the output from the deep learning network for the first photo of Unit 1, while P_1 represents the average output across all photos of Unit 1.

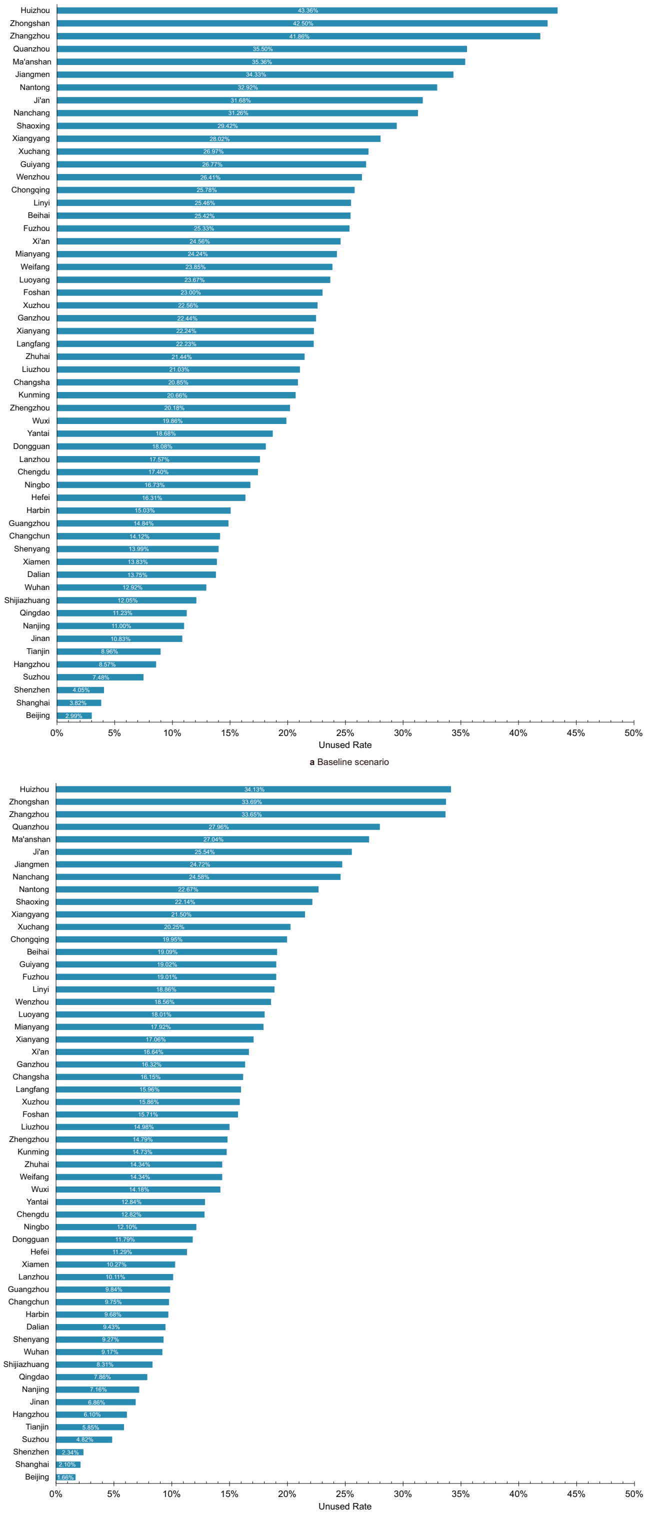

Fig. 2 | City-level unused-housing rates. The figure in (**a**) displays the distribution of city-level stock-based unused rates (SURs) in the baseline scenario in early 2021. The figure in (**b**) displays the distribution of city-level SURs in the conservative scenario, reproduced based on (**a**).

2001 and 2018, at least one unit had never been occupied by early 2021. Based on the volume of housing completions in real estate development, the unused rate of 17.4% can be converted to a total volume of 0.93 billion m^2 of unused housing in these 56 cities. Two facts may facilitate an intuitive comprehension of this magnitude. First, the unused volume is equivalent to 293% of the annual housing completions in these 56 cities in 2020. Second, the unused volume can accommodate 24.19 million residents, or 6.0% of the urban population in these 56 cities, based on the per-capita living space for urban Chinese residents of 38.6 m^2 in 2020.

Besides the baseline estimate in which we adopt the most likely parameters, we also provide a conservative estimate that tends to achieve the lower bound of SUR (detailed in "Methods"). The conservative estimate, as illustrated in Fig. 2b, puts the unused rate and volume at 12.5% and 0.67 billion m^2, respectively—a lower but still striking number.

Figure 2 also demonstrates substantial inter-city variances, particularly from an across-tier viewpoint. Among the 56 sample cities, there are 4 first-tier cities (7.1%), 18 second-tier cities (32.1%), and 34 third-tier cities (60.7%). The detailed definitions of three tiers of cities are presented in Supplementary Table 1. In three first-tier cities—Beijing (3.0%), Shanghai (3.8%), and Shenzhen (4.1%)—the SUR is below 5%. Guangzhou is the only first-tier city with a double-digit SUR (14.8%). The second-tier cities witness larger variations: the SUR is moderate in cities such as Suzhou (7.5%), Hangzhou (8.6%), and Tianjin (9.0%), but remarkable in several other second-tier cities in West China, such as Chengdu (17.4%), Xi'an (24.6%), and Chongqing (25.8%). Most third-tier cities have a large proportion of unused housing, with a weighted average SUR of 25.3%. Specifically, the unused rate exceeds 30% in nine third-tier cities. Although we leave more conclusive analyses on the interpretations of the between-tier variations to future study, the substantial gap still provides meaningful information to facilitate a back-of-the-envelope calculation on the national-level SUR. Given that over 250 cities not included in our sample are all third-tier cities, interpreting the weighted-average SUR of the 56 sample cities (17.4%) as the lower bound of the national-level SUR is plausible. In this case, the total volume of unused housing at the national level would reach 1.76 billion m^2 in early 2021. Here, we can provide another benchmark for comparison. According to the official statistics by the National Bureau of Statistics of China, the national-level volume of developers' housing inventory was 223.79 million m^2 by the end of 2020[17], which implies the volume of unused units held by households was almost eight times that of the well-known housing inventory held by developers.

Figure 3 presents evidence that such an inter-city variation pattern is consistent with the widespread concern about the potential oversupply in China's housing market. Figure 3a splits the sample cities into three categories based on the ratio between aggregated housing completions in 2001–2020 and population growth during the same period, which serves as a proxy for excess housing supply in the city. The results indicate that the unused rate was significantly higher in cities with a larger supply-demand ratio during these two decades. In the next two panels, we divide the sample cities according to two major housing-supply determinants disclosed by the literature. Geographically, Fig. 3b reveals the SUR was significantly higher in cities with higher land-supply elasticity, measured by the quantity of flat land area (i.e., area of non-water land with a slope below 15 degrees) in the city, normalized by the population in 2000[19]. Institutionally, Fig. 3c demonstrates the SUR was significantly higher in cities with larger budget deficits in 2001–2020, which serves as a proxy for local governments' dependence on income from residential land sales as off-budget revenues[20]. In other words, a city is more likely to witness a higher SUR if it has more developable land resources for housing development and/or if its local authority has to sell more residential land to generate off-budget revenues, which further attests to the linkage between the high unused rate and potential housing oversupply.

The within-city analysis, as depicted in Fig. 4, also supports such a linkage. We depict the within-city distribution of all 56 sample cities in Supplementary Note 2, which indicates that, within the same city, the

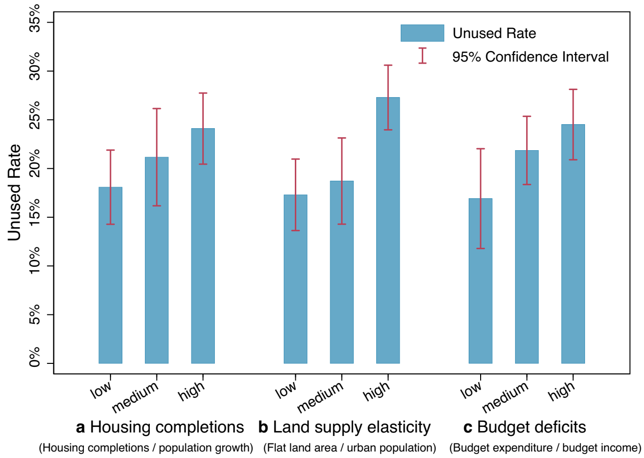

Fig. 3 | Relationship between housing oversupply and unused rates. The blue bars indicate the mean value of city-level stock-based unused rates (SURs) for the low, medium, and high groups, divided based on the tertiles in terms of **a** the ratio between aggregated housing completions in 2001–2020 and population growth during the same period, which indicates excess housing supply, **b** the quantity of flat land area in the city normalized by the population in 2000, which measures the land-supply elasticity, and **c** the government budget deficits in 2001–2020, which measures local governments' incentives to sell residential land. For all the three panels, the sample size is 18 for the low group, and 19 for the medium and high groups. The blue bars represent the estimated value of unused rate for each group. The error bars in red depict 95% confidence intervals.

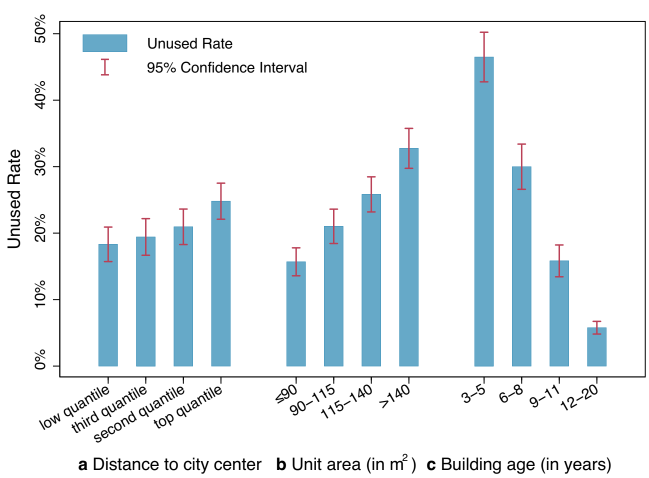

Fig. 4 | Within-city variances of the unused rates. a Indicates the mean value of the community-level stock-based unused rates (SURs) for the four groups, divided based on the quantiles of distances to the city center, respectively. **b** Indicates the mean value of the unit-level SURs for the groups with unit areas of no more than 90 square meters (m^2), between 90 and 115 m^2, between 115 and 140 m^2, and more than 140 m^2, respectively. **c** Indicates the mean value of the unit-level SURs for the groups with building years aged 3–5 years, 6–8 years, 9–11 years, and 12–20 years, respectively. The sample size is 56 for each of the 12 groups of the three panels. The blue bars represent the estimated value of unused rate for each group. The error bars in red depict 95% confidence intervals.

unused housing phenomenon spreads widely across communities, instead of concentrating in a few neighborhoods. However, the unused rate is still significantly higher in the suburbs (i.e., communities whose distances to the city center are above the top quartile; Fig. 4a), which is consistent with the pattern that housing oversupply is more likely to emerge in the suburbs in contemporary China[20]. Similarly, larger dwelling units tend to have a higher unused rate due to potential oversupply (over 140 m^2 in unit size; Fig. 4b). In addition, SUR tends to decrease with building age (Fig. 4c): the unused rate in the building cohort aged 3–5 years reached 46.5% and then dropped to 30.0% in the cohort of 6–8 years, 15.8% in the cohort of 9–11 years, and 5.8% in the cohort of 12–20 years. On average, the unused units had remained unused for 6.5 years by early 2021. Nevertheless, we provide the analysis of the "constant-quality" SURs (Supplementary Note 3), which controls for the effect of micro-level housing attributes on the city-level unused rates. The results demonstrate the composition effect within the cities does not drive the patterns shown in Figs. 2 and 3.

Effect on carbon emissions

We then convert the estimated volume of unused housing to carbon emissions. As detailed in "Methods," for each square meter of unused housing in early 2021, we calculate its annually amortized materialization carbon emissions (with an expected service lifespan of 50 years) and annual central-heating carbon emissions (only for cities in Northern China), respectively, assuming unused housing does not generate other operating carbon emissions, such as cooking and lighting. We can then calculate the aggregated volume of avoidable carbon emissions associated with all the unused dwelling units in 2020. As shown in Fig. 5, the total volume of preventable carbon emissions associated with unused housing in these 56 cities amounted to 25.04 million tons of CO_2 in 2020 based on the baseline estimate. Here, we compare the total magnitude with two standards. First, our estimates indicate the total carbon emissions in the residential sector of these 56 cities, including those embedded in construction and materials in new housing and those produced in operating housing, was approximately 420.76 million tons of CO_2 in 2020; thus, unused housing accounted for 6.0% of the total residential carbon emissions. Second, over the past two decades, the Chinese government has prioritized four key measures for decarbonizing the urban residential sector: lowering carbon intensity in steel and cement production, promoting prefabricated buildings with lower carbon emissions in the construction stage, promoting green buildings with lower carbon emissions in the operation stage, and renovating existing buildings to improve energy efficiency[2,6,9]. We estimate that, compared with the counterfactual scenario where the carbon emission intensity remains at the level of the year 2000, these four measures achieved a carbon emission reduction of 107.91 million tons of CO_2 in the residential sector of these 56 cities in 2020. In other words, the unnecessary carbon emissions of unused housing were equivalent to 23.2% of the carbon emission reductions due to these four measures. Therefore, excessive housing supply and the massive volume of unused housing have significantly impeded China's continuous efforts to decarbonize the residential sector. We also present a highly conservative estimate, which not only adopts the conservative estimate of the volume of unused housing as mentioned in the last section, but also assumes 30% of the central-heating system in the unused dwelling units would be turned off (detailed in "Methods"). As depicted in Fig. 6, unused housing's carbon emissions would reduce to 17.20 million tons in 2020 if we adopted this conservative estimate, which still accounts for 4.1% of overall residential carbon emissions and offsets 16.2% of carbon emission mitigation in the 56 sample cities.

We also estimate the overall volume of preventable carbon emissions associated with unused housing throughout the country. Specifically, we adopt the weighted-average unused rate of the 56 sample cities (17.4%) as the national-level housing unused rate to consider the other 250 cities. In this case, unused housing generated 55.81 million tons of CO_2 in 2020 based on the baseline estimate. The overall magnitude accounts for 6.9% of the total residential carbon emissions and offsets 26.5% of carbon emission mitigation at the national level. If we adopted the conservative scenario, the unused housing's carbon emissions would reduce to 39.85 million tons in 2020, still accounting for 4.9% of the total residential carbon emissions and offsetting 19.2% of carbon emission mitigation.

https://doi.org/10.1038/s41467-025-57217-7

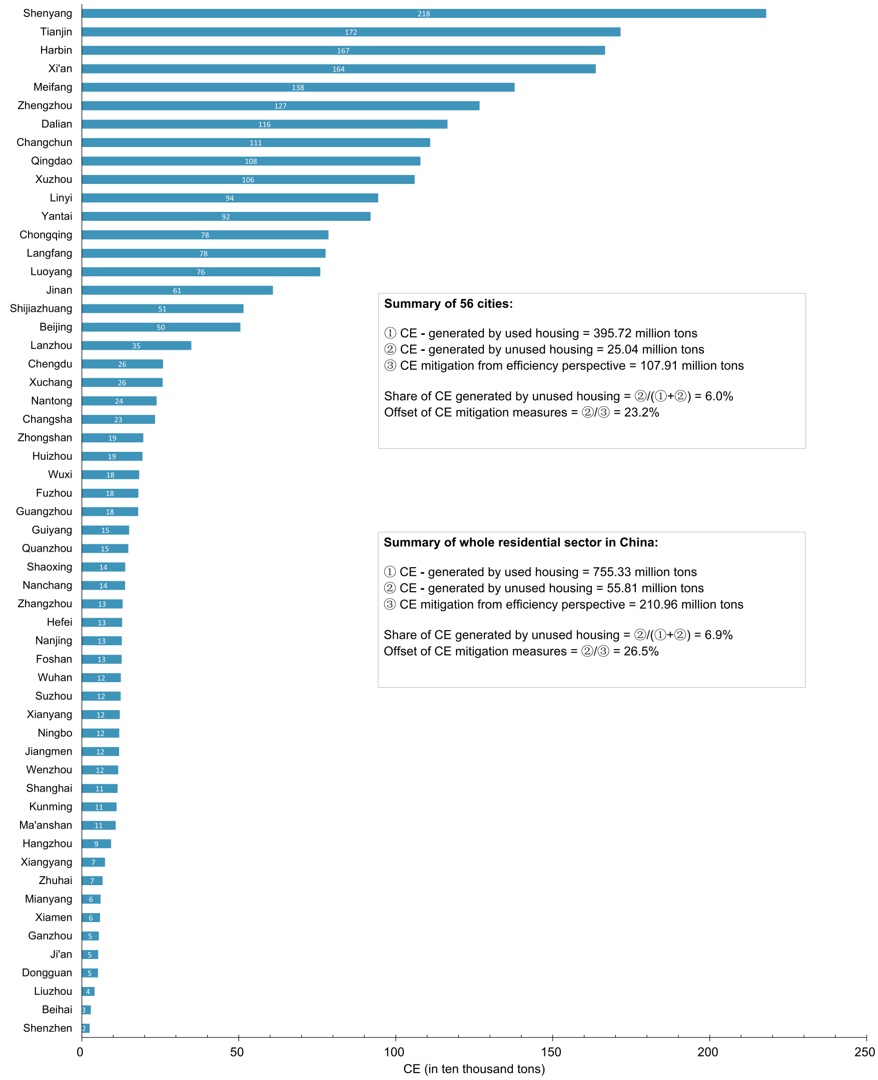

Fig. 5 | Impact of unused housing on carbon emissions in 2020. This figure depicts the city-level carbon emissions (CE) generated by unused housing in 2020. The 56 cities in our sample are sorted by the total carbon emissions of unused housing in 2020.

In Supplementary Note 4, we also provide two other methods to estimate the unused housing carbon emissions in 2020. The results are close to the current method, although the alternative methods must rely on more assumptions.

Consistent with Fig. 2, the results in Fig. 5 emphasize the importance of unused housing in second- and third-tier cities. For example, in Xi'an, the capital city of Shaanxi Province, with an estimated SUR of 24.5%, the carbon emissions generated by the construction and operation of unused housing amounted to 1.64 million tons, which accounted for 13.3% of its residential carbon emissions (12.30 million tons) in 2020. By contrast, in superstar cities with low unused rates, such as Beijing, the impact of unused housing on carbon emissions is modest. Meanwhile, unused housing's carbon emissions are generally higher in Northern China due to central-heating emissions, further increasing the waste associated with unused housing.

Discussion

This study echoes recent literature that highlights the importance of the sufficiency perspective in decarbonizing the residential sector. Using the method based on indoor photos of online listings, our calculations in 56 major Chinese cities indicate 17.4% of dwelling units completed between 2001 and 2018 had never been occupied by early 2021, with the phenomenon being most pronounced in relatively smaller cities that have arguably experienced a massive housing oversupply over the past two decades. The construction and operation of this vast amount of unused housing produce 6.9% of the Chinese residential sector's carbon emissions, which could and should be avoided. The avoidable emissions of unused housing significantly offset China's ongoing efforts to mitigate the residential sector's carbon emissions.

The findings highlight policy priorities in China's subsequent efforts to establish a carbon-neutral residential sector. On the one hand, policymakers should aim to avoid further housing oversupply, primarily through guiding the residential industry by implementing long-term and annual housing development plans, as well as associated residential land-supply schemes, based on high-quality housing-demand forecasts. Lowering local governments' reliance on land-sales revenues and the resulting residential land oversupply through a reform of the current local fiscal system can also contribute to this goal. On the other hand, local governments can also commit to utilizing at least a portion of the unused dwelling units to meet new housing demand, hence partially replacing the demand for new housing construction. This objective can be facilitated by, for instance, imposing a property tax on vacant units, which can increase the homeownership cost for unused dwelling units while minimizing the potential unintended effect on housing demand.

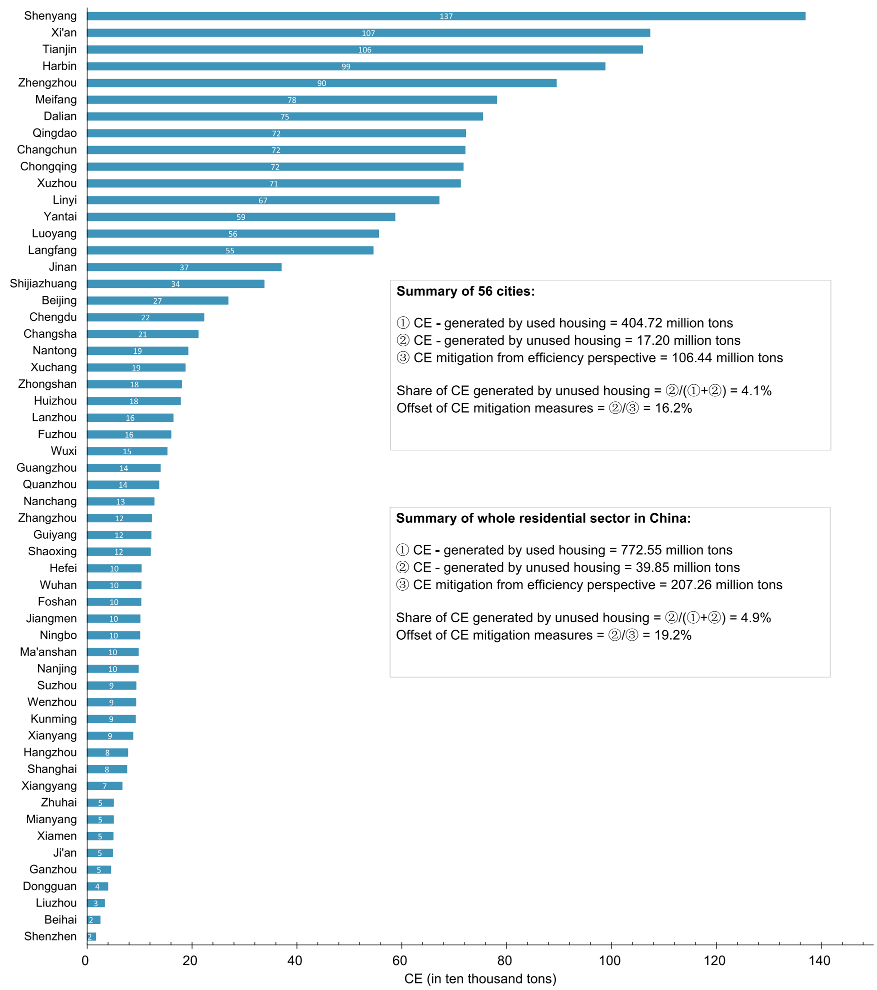

Fig. 6 | Impact of unused housing on carbon emissions in 2020. This figure depicts the city-level carbon emissions (CE) generated by unused housing in 2020 in the conservative scenario.

Here, we provide an intuitive understanding of the potential effect of these policy efforts via the scenario analysis. In the benchmark scenario, we assume the Chinese government makes no attempt to eliminate the increase in unused housing or utilize existing unused housing, and hence, the unused rate at the end of 2030 remains at the same level as at the end of 2020. In the other scenarios, we assume a reduction in the unused rate to 75%, 50%, and 25% as the benchmark level to reflect different policy intensities. Figure 7 presents the impact of these initiatives on China's national-level residential-sector carbon emissions between 2021 and 2030. Compared with this benchmark scenario, the total carbon emissions in the residential sector can be reduced by 4.7% if the unused rate can be gradually lowered to 13.0% in 2030 (i.e., 75% of 17.4%), 9.1% if the unused rate can be lowered to 8.7% (i.e., cut by half) in 2030, or 13.1% if the unused rate can reach 4.3% (i.e., 25% of 17.4%) in 2030. Not surprisingly, the potential achievement of unused-housing-related decarbonization measures is particularly pronounced in cities with current high unused rates, such as Xi'an. We estimate Xi'an's residential carbon emissions in 2021–2030 can be reduced by 11.6% if its unused rate can be cut by nearly half from 24.6% at the end of 2020 to 12.4% in 2030. Besides the aforementioned measures to cut the unused rates, policymakers can also seek to reduce the carbon emissions associated with unused units. In particular, in the northern cities, both technique and policy measures can be implemented to minimize central-heating emissions of the unused units. We encourage researchers and policymakers from other countries, especially developing countries, with massive new home constructions, to consider unused housing when designing decarbonization strategies.

We acknowledge that our research still has a few limitations. First, the conversion from the LUR to the SUR relies on several assumptions, which might affect the accuracy of the SUR estimate. Second, the reliability of our unused-rate estimate could be substantially enhanced if we could introduce some in-site survey data directly measuring the community- or even city-level unused rates. Third, the carbon emission calculations might overlook some components or meaningful between-city variations in carbon emission intensity. Finally, and perhaps most importantly, the current study focuses on the impact of unused housing units on national-level carbon emissions, and hence is not able to shed more light on the inter- and within-city analyses. In particular, the significant differences and dynamics in SUR between different tiers and regions are noteworthy, which would also affect our understanding of the impact of SUR on carbon emissions. More analyses of the inter- and within-city spatial distribution of the unused rate, including its determinants and consequences, should be at the top of the future research agenda.

Methods

Data

Typically, the seller of a listed dwelling unit in China, or their agent, would upload the listing information onto online listing platforms for free to disseminate information to potential buyers[21]. In almost all cases, in addition to the text-format information, the seller or their

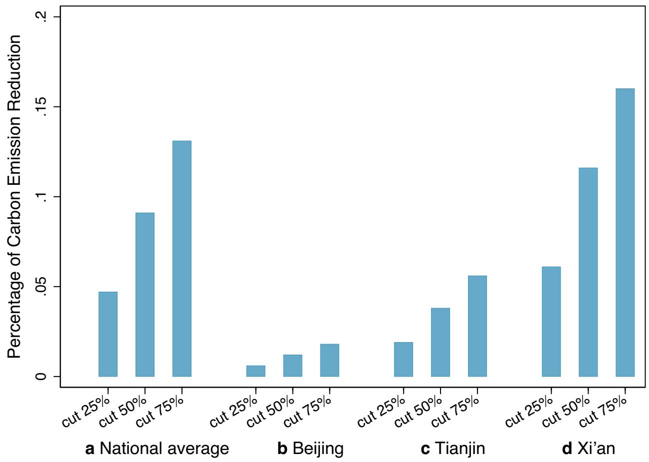

Fig. 7 | Contribution of unused rate decreasing on carbon emission reductions between 2021 and 2030. This figure indicates the percentage of carbon emission reduction in the Chinese residential sector after the unused rate is gradually cut by 25%, 50%, and 75% throughout 2021 and 2030, respectively. In the benchmark scenario, the unused rate at the end of 2030 remains at the same level as at the end of 2020 (for the national average, we adopt the weighted average of 56 sample cities). We make a forecast from both the national level (**a**) and for three representative cities: **b** Beijing, which represents the cities with low unused rates; **c** Tianjin, which represents the cities with medium unused rates; and **d** Xi'an, which represents the cities with high unused rates.

agent uploads multiple indoor photos of the property to depict its current condition, which has proven crucial for attracting buyers' attention[22,23]. Figure 8 presents representative examples of such indoor photos. Based on these examples, we, or a trained deep learning model, can discriminate between unused units (Types I–III in the figure) and units that are currently or were once occupied ("occupied units" hereafter; Types IV and V).

Following this strategy, the raw data are collected from one of the leading online housing listing platforms in China. Note that although the literature points out that agents may intentionally release fraudulent or duplicate listings on online platforms to attract attention[21], this possibility is not a concern for this online platform. As the core competitiveness of this platform, the platform not only manually verifies the validity and uniqueness of each listing, but also provides an explicit guarantee to buyers that they can get monetary reparation from the platform if they find invalid (including fake, duplicate, and outdated) listing information on it.

Through web spiders, we collected information on the dwelling units listed on the platform in 56 major cities around China between October 2020 and August 2021. We then cleaned the sample using the following procedures: (1) We only kept units in communities completed between 2001 and 2018, because we only consider dwelling units that have been unused for at least two consecutive years after completion; (2) we only kept units in communities in the major districts/counties in the city as listed in Supplementary Table 1; (3) we only kept units with at least one indoor photo; (4) we dropped outliers as the top and bottom 1% listed units in the number of households in the community and unit size, and the top 1% listed units in the number of floors; and (5) for each city, we also dropped the top and bottom 1% listed units in the unit price. Finally, the working sample includes 1,196,585 listings.

For each unit, we collected all the information provided on the listing webpage (see, e.g., Supplementary Fig. 2), including the listing date and price, major community-level attributes (e.g., completion year, address, whether equipped with the central-heating system, etc.), major unit-level attributes (e.g., unit size, time of construction, floor, etc.), text descriptions, and, most importantly, all the indoor photos uploaded. We collected 6,577,579 photos for the working sample, giving us an average of 5.5 photos per listed unit. We also list in Supplementary Table 1 the number of dwelling units and photos included in each city.

Notably, we also collected the date of the previous transaction of the listed unit from the platform. In China, the transaction tax rate for a housing resale transaction is partially determined by the seller's holding period, defined as the interval between the date when the seller purchased the unit and the date of the current resale. Therefore, the date of the previous transaction is widely perceived as a crucial attribute of a listed unit and is prominently displayed on the listing webpage.

We provide evidence on the reliability and representativeness of the listing data in Supplementary Note 1.

Identifying unused dwelling units

We randomly chose 40,000 units with 233,896 indoor photos from the working sample and hired three research assistants to label whether each image belonged to an unused or occupied unit. For the research assistants' reference, we provided sample images of unused and occupied units and text descriptions of unused rooms to help them classify each photo of the training set into unused or occupied units. We also examined the photos for which the research assistants had divergent opinions.

As shown in Supplementary Table 2, among all 233,896 photos in the training set, 38,322 photos are identified as unused, and the other 195,574 photos are identified as occupied. In particular, of the 38,322 photos classified as unused, 23,364 have no decoration (Type I), 12,437 are partially decorated (Type II), and only 2521 are well decorated (Type III), which indicates the majority of unused units have no or at most partial decoration and, thus, are relatively easy to distinguish from occupied units. Nevertheless, we did not directly use this original training set, because the ratio between unused and occupied classes is about 1:5, and the existing literature points out that classifiers trained on such a class-imbalanced dataset tend to be overwhelmed by the larger class and ignore the smaller one[24]. Specifically, the trained classifiers are inclined to categorize all observations into the majority class or, equivalently, reduce the recall rate of the minority class. In our context, if the trained classifier directly classifies all photos into the occupied group, the model would still achieve a high accuracy rate due to the highly unbalanced sample distribution. However, such a classification is obviously ineffective. To solve the overfitting problem caused by an imbalanced training set, we utilized the undersampling method, a widely recognized strategy in the literature[25,26], and randomly selected 38,322 photos from the occupied group to ensure that the size of the two classes adopted in the following training is balanced.

We use ResNet-50, a deep learning network widely used for image classification[27–29], as our backbone network. Different from convolutional networks such as VGG16, ResNet-50 reformulates the layers by learning residual functions with reference to the layer inputs, instead of learning unreferenced functions, enabling the network to be deeper and achieve higher classification accuracy[30]. The architecture of the network we used is depicted in Supplementary Fig. 3. We convert the classification network into a regression network, whose continuous output is more applicable when we comprehensively consider the outputs of all photos belonging to the same listed unit. More specifically, we use a one-way fully connected layer with sigmoid as the output layer to replace the original 1000-way fully connected layer with softmax. In the current setting, our network returns a value ranging from 0 to 1, representing the likelihood that the input photo belongs to an unused unit. The key component of our network is the convolutional kernel (illustrated by "Conv" in Supplementary Fig. 3), which is essentially an n×n weighting matrix that extracts features from the outputs of the last layer. In other words, the sequence of stacked convolutional layers can be regarded as a more intricate and sophisticated feature extractor, which transforms the input red-green-blue

Fig. 8 | Examples of photos for unused and occupied dwelling units. This figure displays representative indoor photos from the online housing listing platform, which are publicly available under their terms of use and permitted for non-commercial academic purposes (see https://gitee.com/hefanz/unused-housing-in-urban-china-and-its-carbon-emission-impact/blob/master/PLATFORM%20SERVICE%20USE%20AGREEMENT.pdf for details). Any identifying information has been removed to ensure privacy. The rooms in Type I belong to units without interior decoration, whereas the rooms in Type II are partially decorated; obviously, neither of these two types of units meets the ordinary living standard. The Type III rooms are well decorated, but their pristine state and brand-new furniture indicate they have never been inhabited. Type IV rooms, on the other hand, are filled with an abundance of daily items, showing they are currently occupied. Type V rooms are nearly empty; nevertheless, we can still infer the rooms were once occupied based on the traces of usage on the walls, floors, and ceilings, even if the owners/renters have moved out. Accordingly, we classify the dwelling units in Types I, II, and III as "unused," and units in Types IV and V as "occupied".

(RGB) image into dozens of features that can be comprehended by subsequent layers. The training process is analogous to using the training sample to teach the network which features should be extracted and the relationships between features and regression results. Besides the original labeled images, we also generate random horizontal reflections to modify the RGB channel intensities of the input training images to prevent overfitting issues[31].

To accelerate training, we start with the pre-trained parameters of ResNet50, which was trained using IMAGENET, a dataset containing over 1 million images. Then, we fine-tune the model based on our dataset of tagged photos. We randomly choose 90% of the tagged photos as the training set and the remaining 10% as the testing set. Using a batch size of 96 and a learning rate that initializes at 0.002 and decays by 0.9 per 10 epochs, we train the model with an Adam optimizer for 1000 epochs.

Because the output result is a continuous possibility, we classify the photo based on a designated threshold. Specifically, if the output result is larger than the designated threshold, the photo is classified as positive (i.e., unused). Similarly, if the output result is smaller than the designated threshold, the photo is classified as negative (i.e., occupied). To select an optimized threshold, we use the F1 score, a commonly used metric for evaluating the performance of deep learning models[32,33], as the evaluation metric. As listed in Eq. (1) and Eq. (2), True Positives (*TP*) are examples correctly classified as positives, False Positives (*FP*) refer to negative examples incorrectly classified as positives, True Negatives (*TN*) correspond to negative examples correctly labeled as negative, and False Negatives (*FN*) refer to positive examples incorrectly labeled as negatives. *P*(*Precision*) reflects the capacity of a classification model to identify only relevant data, whereas *R*(*Recall*) reflects the ability to identify all relevant cases within a dataset[34]. The F1 score, as listed in Eq. (3), is the harmonic mean of *P* and *R*, allowing it to assess the model comprehensively:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{3}$$

Supplementary Fig. 4 depicts the three metrics at various thresholds. When the threshold is 0.44, we get the highest F1 score of 0.897, with a *P* of 90.1% and a *R* of 89.3%. For simplicity, in the baseline estimate, we directly apply the threshold of 0.5 to distinguish between the unused and occupied classes. Under this threshold, *P* is 91.4% and *R* is 87.5%.

We send each of the 6,577,579 photos into the trained model. The mean value of the predicted likelihood is 0.24, whereas the standard deviation is 0.34. As Supplementary Fig. 5a shows, the distribution of prediction concentrates around 0 (i.e., 100% to be occupied) and 1 (i.e., 100% to be unused); specifically, 88% of predicted results are smaller than 0.2 or larger than 0.8. Among all 6,577,579 photos, 1,284,981 (19.5%) have a prediction value larger than the threshold of 0.5 and hence are classified as unused.

For each of the 1,196,585 listed units, we calculate the average prediction value of all the photos associated with the unit as its unit-level prediction value. The unit-level prediction has a mean value of 0.23 and a standard deviation of 0.31. As Supplementary Fig. 5b demonstrates, the unit-level predictions are still centered around 0 and 1, with 89% of units possessing a prediction value smaller than 0.2 or larger than 0.8. Given the threshold of 0.5, we find 211,498 (17.7%) units are identified as unused, while the other 985,087 are identified as occupied.

We provide evidence of the reliability of the LUR in Supplementary Note 1.

Converting LUR to SUR

The key challenge in converting LUR to SUR lies in the potential difference in the selling probability between unused and occupied units. For instance, if owners are less willing to list and sell unused units, the LUR would be systematically lower than the SUR, and vice versa.

Suppose S dwelling units exist in the housing stock of city X at time T; ρ percent of the stock is still unused (i.e., the stock-based unused rate, or SUR), and $1-\rho$ has been occupied. Here, we focus on adjusting for two potential differences between the unused and occupied groups. First, the selling probability of an unused unit might not be equal to the selling probability of an occupied unit. We assume that the selling probability of occupied units during our sample period is p and that of unused units is qp. Second, the extent to which these two groups rely on online platform listings may also differ. We assume that δ of occupied units for sale are listed on our online platform, whereas the corresponding ratio for unused units is $r\delta$. Accordingly, LUR can be calculated as Eq. (4):

$$LUR = \frac{S \times SUR \times qp \times r\delta}{S \times (1 - SUR) \times p \times \delta + S \times SUR \times qp \times r\delta} \tag{4}$$

Then, we can have Eq. (5):

$$SUR = \frac{LUR}{LUR - LUR \times q \times r + q \times r} \tag{5}$$

We estimate q, the ratio between the selling probability of unused and occupied groups (i.e., and), based on the transaction record data from the online listing platform. Specifically, for each listed unit, the online platform reports the date of the prior sale of the unit (i.e., when the current owner of the listed unit purchased the unit from the new home or resale market), allowing us to calculate the holding period of the current owner. We can hence calculate the average lengths of holding periods for the unused and occupied groups (i.e., H_Unused and $H\_Occupied$, respectively) in each city. By definition, the average (annual) selling probability of a specific dwelling-unit cohort equals the reciprocal term of the average holding periods (in years) of the same cohort. Accordingly, q can be calculated as Eq. (6):

$$q = \frac{P_{Unused}}{P_{Occupied}} = \frac{\frac{1}{H\_Unused}}{\frac{1}{H\_Occupied}} = \frac{H\_Occupied}{H\_Unused} \tag{6}$$

We assume r, the difference in the likelihood of listed units appearing on the online platform, equals 1 in the baseline estimate based on the following reasons. First, the China Institute of Real Estate Appraisers and Agents (the Chinese counterpart of the National Association of Realtors) reports over 85% of housing resale transactions in China are assisted by professional agents, who, in almost all cases, rely heavily on online platforms to disseminate information. Therefore, one can assume a very high proportion of listings would appear on online platforms, leaving little potential for differences between unused and occupied groups. Second, existing studies in both the United States and China indicate occupancy status is not a key determinant of whether sellers choose an agent service or online listing service[35,36]. Third, to test whether the online platform contains significant sampling biases, we collected listing data for five sample cities between November 2020 and March 2021 from both our online platform and another leading online listing platform in China, and then applied the same classification procedures to calculate the city-level LURs for each platform. As Supplementary Table 3 shows, the LURs based on these two platforms are highly consistent; in particular, the LURs for our platform are neither systematically higher nor lower than those based on the other platform. We can thus safely assume that the online platform we choose neither oversamples nor undersamples unused units across all online listings.

In the calculation, because Fig. 4 indicates that LUR substantially varies with building age, we split the housing communities in a city into four groups according to the year of construction completion. The first three groups include housing communities completed in 2016–2018 (i.e., with buildings aged 3–5 years), 2013–2015 (6–8 years), and 2010–2012 (9–11 years), respectively, whereas the last group comprises communities completed in 2001–2009 (12–20 years). For each city, we first calculate the LUR and q for each building-age cohort. Then, we calculate the SUR for each building-age cohort based on Eq. (5). Finally, we calculate the weighted average of SUR for all four building-age cohorts, weighted by the volume of housing completions in the corresponding years in the city as reported by the local statistical authority. In Supplementary Note 1, we also try grouping the units by other housing attributes, such as location, instead of building age, and the results are qualitatively constant.

Conservative estimate of the unused rate

We expect the baseline estimate based on the aforementioned parameters to achieve an estimate of the most probable SUR. We also evaluate a conservative estimate, which we expect to yield a lower bound of SUR. Compared with the baseline estimate, the conservative estimate alters two key parameters. First, in classifying the unused and occupied units based on indoor photos, instead of adopting the conventional threshold of 0.5, we use the $F_{0.5}$ score as the metric to get the threshold, which values P(*Precision*) more than R(*Recall*). The $F_{0.5}$ score is calculated as Eq. (7):

$$F_{0.5} = (1 + 0.5^2) \times \frac{P \times R}{0.5^2 \times P + R} \tag{7}$$

We obtain the maximum $F_{0.5}$ score of 0.922 when the threshold is 0.7, with P improving from 91.4% to 95.6% and R decreasing from 87.5% to 80.7%. A higher P ensures the model is less likely to misclassify occupied units to the unused class. Under the conservative threshold of 0.7, the number of unused units is reduced to 180,600 from 211,497 in the baseline estimate.

Second, instead of assuming r equals 1 when converting LUR to SUR, we consider the possibility that unused units may have a higher probability of being listed on online platforms. Specifically, as described above, the official statistics indicate that 85% of housing resales are assisted by professional agents. As a most extreme case, we assume 100% of unused listed units are assisted by agents and therefore appear on online platforms. In this scenario, for each sample city, we can impute the share of occupied units assisted by agents (and hence listed online) based on the estimated LUR, bringing to 85% the weighted average of being assisted by agents of both groups. We then use the imputed r, instead of the value of 1, in Eq. (5) to convert LUR to SUR.

Basic setting of carbon emission calculations

For city i, between 2001 and 2020, we can observe the annual series of housing completions (in floor area), $AC_{i,t}$. Meanwhile, we can impute the annual series of total housing stock (in floor area), $AS_{i,t}$, between 2001 and 2020, as Eq. (8):

$$AS_{i,t} = AS_{i,t-1} + AC_{i,t} - AD_{i,t}, \tag{8}$$

where $AD_{i,t}$ refers to the floor area of housing demolition in the city-year, calculated based on the annual demolition rate of 2% with the expected service lifespan of 50 years as required by the technique code in China[37]. To calculate the housing stock in 2000, we use the city's urban population and per-capita living space, as reported by the 2000 Population Census.

Note that, as revealed in this study, a portion of $AS_{i,t}$ had never been occupied by the end of year t. Here, we assume that the housing stock in 2000 was completely occupied by 2001 due to the inadequate housing supply during the pre-reform era and only consider the unused units completed in and after 2001. In our data, we can directly observe the SUR associated with each building-year cohort in early 2021 (or the end of 2020). Hence, based on $AC_{i,t}$ and the building-year-specific SUR, we can calculate the volume of unused housing in each building-year cohort at the end of 2020 (SUR of the year-2019 cohort is set as 100% by definition), denoted by $AU_{i,2020,BuildingYear}$, from $AU_{i,2020,2001}$ to $AU_{i,2020,2019}$. Note the unused rate is originally calculated based on the proportion of units, and here, we apply it to the floor-area volume. Considering that larger units are more likely to be unused, as revealed in Fig. 4 in the main text, this conversion achieves a lower bound of the estimate on unused volume (and the associated carbon emissions). By aggregating the volume of unused housing in the building-year cohorts between 2001 and 2019, we can get the total volume of unused housing at the end of 2020, or $AU_{i,2020}$, as Eq. (9):

$$AU_{i,2020} = \sum_{Building\ Year=2001}^{2019} AU_{i,2020,Building\ Year}. \tag{9}$$

We can then get the floor area of occupied units, or $AO_{i,2020}$. Note we assume it is reasonable for any housing unit to be unused during the first two years after its completion. In China's common practice, a buyer typically takes several months to over one year to decorate a newly purchased unit and then move in. Here, we choose to adopt a relatively longer option of a two-year window rather than a one-year window to obtain a lower bound of the unused rate. For this purpose, we also need to further extract the housing-completion area in 2020 from the housing-stock area in 2020 to calculate the floor area of occupied units at the end of 2020, as Eq. (10):

$$AO_{i,2020} = AS_{i,2020} - AU_{i,2020} - AC_{i,2020}. \tag{10}$$

$AO_{i,t}$ can be interpreted as the floor area of actual housing demand in the city-year. For each sample city, we can use the logistic function to regress the annual series of $AO_{i,t}$ between 2001 and 2020 and then use the estimated coefficients to forecast $AO_{i,t}$ between 2021 and 2030.

Calculations of actual carbon emissions in 2020

The residential sector's overall carbon emissions consist of the construction and operation stages. The volume of residential-building construction emissions in year t, $CE\_C\_Actual_{i,t}$, is calculated as Eq. (11):

$$CE\_C\_Actual_{i,t} = AC_{i,t} \times CO2\_C_{i,t}, \tag{11}$$

where $CO2\_C_{i,t}$ refers to the embodied carbon-intensity factor during the construction stage (including the construction material productions, e.g., cement and steel) for the residential sector in year t.

For the operation stage, the volume of carbon emissions of the residential building stock in year t, $CE\_O\_Actual_{i,t}$, is calculated as Eq. (12):

$$\begin{aligned} CE\_O\_Actual_{i,t} = AO_{i,t} \times (CO2\_O_{i,t} + CO2\_CH_{i,t}) \\ + AU_{i,t} \times CH_i \times CO2\_CH_{i,t}, \end{aligned} \tag{12}$$

where $CO2\_O_{i,t}$ indicates the operational carbon-intensity factor (e.g., cooking and lighting) in China's residential sector. For cities located in Northern China, we consider the carbon emissions associated with central heating in different regions. Because the regional energy structure and climate conditions greatly influence heating demand and carbon emissions, we employ a provincial-level factor, $CO2\_CH_{i,t}$, to accurately measure the carbon emission intensity of central heating[38]. We do not consider the carbon emissions associated with the cooling system, because the district cooling system has seldom been adopted in China's residential sector so far[39,40].

Here, we assume that an unused unit does not generate operational carbon emissions. As for the heating emissions, we assume that if an unused unit uses the individual household-based heating system, the owner will turn off the system, and hence, the unit would no longer generate heating emissions. However, for dwelling buildings with central-heating systems, whose proportion in Northern Chinese cities reaches as high as 89.5% according to our dataset, the heating systems are expected to keep functioning for the unused dwelling units for two reasons. First, from the community perspective, a heat-supply company is typically reluctant to turn off the central-heating system of the whole community even if the community is underoccupied. Chinese governments have implemented official regulations on operating the central-heating system for residential communities. Normally, the heating facilities shall keep functioning during their warranty period (typically the first two years after completion) for all communities with viable central-heating systems, regardless of the occupancy conditions. Post-warranty, these systems are required to remain operational if a desirable proportion of households (usually 40–50%) have moved in and continue to use central heating. Even if the occupancy rate falls slightly below the official threshold, in practice, the heat-supply company still tends to continue supplying heat to avoid dissatisfaction among remaining residents. Second, from the perspective of individual housing units, owners of some unused units may have access to turning off their indoor central-heating system. Still, we can reasonably expect that they have limited economic incentives to do so, because the primary charging method for central heating in current China is still based on heating area rather than actual heating consumption[41,42]. Even for the latter case, ~15–30% of the heating expenditures shall still be paid as an infrastructure operation fee even if the indoor heating system is turned off. Based on the above analysis, we only adopt CH_i to indicate the proportion of unused housing with central heating in city i, calculated based on the same dataset of listing-unit information that we use to calculate LUR.

Nevertheless, to account for the potential shutoff of the central-heating system in unused housing, we also introduce a conservative scenario where we assume 30% of the unused units would have their central-heating system turned off either by the community or their owners. In this case, the unused units whose central-heating systems have been shut off still cannot save all the heating emissions for three reasons. First, if the heat-supply company discontinues its service to an underused community, the remaining residents might resort to inefficient self-heating methods (e.g., air conditioners), potentially resulting in higher carbon emissions. Second, a large amount of heat has already been lost before it reaches the final end of housing units. Specifically, the power consumption for the circulating water pump, heat-source efficiency, and heat loss in the pipe network can account for 25–47% of the total heating consumptions[43], which indicates the end-use heat reductions can only have a limited impact on the total heat loss. Finally, households will require more heat to keep their rooms warm once their neighbors turn off their heat supply. The intuition is that the indoor temperature of a single unused housing unit within a large apartment building would not drop significantly after turning off its own heat supply, because its adjacent rooms would heat it up through the walls, floors, and ceilings, with the cost being that the heating systems of these adjacent rooms have to generate more heat. Specifically, the existing literature indicates the long-term closure of heating in an adjacent unit would increase heating expenses by 20–50%[44,45]. Considering all these factors, we assume 30% of the heating emissions from unused housing will not be eliminated merely by shutting down the central-heating system.

As for the three carbon-intensity factors ($CO2\_C_{i,t}$, $CO2\_O_{i,t}$, and $CO2\_CH_{i,t}$), to reflect the actual emissions condition, we have already taken into consideration China's decarbonization efforts in the residential sector. Specifically, we adopt 465.59 kg m^{-2} for $CO2\_C_{i,2000}$, 28.21 kg m^{-2} for $CO2\_O_{i,2000}$, and 124.00 kg m^{-2} for the national-level $CO2\_CH_{i,2000}$[38,46–54] for the scenario in 2000, under which no decarbonization measures had been taken. Then, we consider four key decarbonization measures during the past two decades to update those intensity factors in each subsequent year from 2001 to 2020.

For the embodied carbon emission factor, $CO2\_C_{i,t}$, we first consider reducing carbon intensity in major material productions, specifically, steel and cement. We incorporate the estimates on the proportion[48] of brick concrete structures, steel concrete structures, and steel structures in the national-level annual residential building completions and the average material intensity[46] of 10 major materials (e.g., cement, steel, etc.) for each of these three structures. We then use the time-variant carbon-emission-intensity factors of cement[51] and steel[52–54] and the time-invariant intensity factors of the other eight materials, as well as transportation and on-site construction[47] from existing literature. The average embodied carbon intensity can thus be calculated. Second, we consider the development of prefabricated buildings. The government report documents that the proportion of prefabricated buildings in the national-level residential building completions increased from 4.9% in 2016 to 20.5% in 2020. Prefabrication offers a 15% carbon reduction relative to the traditional cast-in-situ method[55,56]. Consequently, the average embodied carbon intensity is further reduced.

For the operational and central-heating carbon emission factors, $CO2\_O_{i,t}$, and $CO2\_CH_{i,t}$, we account for improvements in energy-efficiency standards and the promotion of green buildings. China has progressively implemented the building energy-efficiency code since 1986, which has gone through four stages, with the goal of the building energy-saving rate increasing from 30% to up to 75%[57]. In addition to new buildings that have implemented the updated standards, the energy-efficiency level of existing residential buildings has also improved, according to the report of the Chinese government. Using completion and renovation data, we can impute the distribution of housing across different energy-efficiency levels, which allows us to calculate the average operational and heating carbon-intensity factors for each year. Additionally, studies show carbon emissions from green residential buildings are 10% lower than those from non-green buildings[58]. Moreover, the Chinese government has also disclosed the proportion of green buildings in total new construction, mushrooming from 2% in 2012 to 65% in 2019. The yearly operational and heating carbon-intensity factors can thus be obtained considering the joint effect achieved by the above two measures.

Finally, we can have the city-level total emission volume in any specific year. Take the year 2020 as an example, as shown in Eq. (13):

$$CE\_Actual_{i,2020} = CE\_C\_Actual_{i,2020} + CE\_O\_Actual_{i,2020}. \tag{13}$$

Calculations of carbon emissions from unused housing

For each square meter of dwelling space that remained unused at the end of 2020 (note we do not include the carbon emissions of unused units completed in 2019 and 2020, because we assume they are not preventable), we consider its related carbon emissions from two perspectives. The carbon emissions for the operation stage are straightforward. As described in the previous subsection, we assume unused housing does not generate operational carbon emissions, but it would still incur central-heating emissions if it is located in a northern city and is equipped with a central-heating system. For the construction stage, we convert the lump-sum construction emissions to the annual amortized emissions[59]. Because the literature has not achieved a consensus on the "discount rate" of carbon emissions, we choose to evenly amortize the construction carbon emissions on the expected service lifespan of 50 years. Because, by definition, the unused period can only exist at the beginning of the lifespan, adopting the 0% discount rate (i.e., even amortization) achieves a lower bound of the annual amortized emissions. Therefore, for dwelling units that were completed in the year *BuildingYear* and remained unused at the end of 2020, the associated waste in carbon emissions in 2020 can be calculated as Eq. (14):

$$\begin{aligned} CE\_Unused_{i,2020} &= \sum_{BuildingYear=2001}^{2018} CE\_Unused_{i,2020,BuildingYear} \\ &= \sum_{BuildingYear=2001}^{2018} (2\% \times AU_{i,2020,BuildingYear} \\ &\quad \times CO2\_C_{i,2020,BuildingYear} + AU_{i,2020,BuildingYear} \\ &\quad \times CH_i \times CO2\_CH_{i,2020,BuildingYear}). \end{aligned} \tag{14}$$

Carbon emission reductions by decarbonization measures

We consider a counterfactual scenario in which none of the four decarbonization initiatives are implemented, and the carbon-intensity factors remain constant at 2000 levels. Under this non-decarbonization scenario, the residential sector's total carbon emissions can be re-calculated as Eq. (15):

$$\begin{aligned} CE\_High_{i,2020} = {} & AC_{i,2020} \times CO2\_C_{i,2000} + AO_{i,2020} \times (CO2\_O_{i,2000} \\ & + CO2\_CH_{i,2000}) + AU_{i,2020} \times CH_i \times CO2\_CH_{i,2000} \end{aligned} \tag{15}$$

The difference between $CE\_High_{i,2020}$ and $CE\_Actual_{i,2020}$ reflects the carbon-emission-reduction contribution of these four decarbonization measures.

Scenario analysis for 2021–2030

Based on $AO_{i,t}$, we can impute the series of annual floor areas of housing completion between 2021 and 2030, $AC\_Base_{i,t}$, which could make the unused rate remain at the same level as the end of 2020. Specifically, recall the SUR is defined as Eq. (16):

$$SUR_{i,t} = SUR_{i,2020} = (AS_{i,t} - AO_{i,t})/AS_{i,t}. \tag{16}$$

Based on Eq. (8) and Eq. (16), we can have Eq. (17):

$$AC\_Base_{i,t} = \{(AS_{i,t-1} - AD_{i,t}) \times (SUR_{i,2020} - 1) + AO_{i,t}\}/(1 - SUR_{i,2020}). \tag{17}$$

We also assume the carbon-intensity factors are consistent with those in 2020. Based on these assumptions, the carbon emission in the baseline scenario can be calculated as Eq. (14):

$$CE_{\_Baseline_i} = \sum_{t=2021}^{2030} \left(AC_{Basei,t} \times CO2\_C_{i,2020} + AO_{i,t} * (CO2\_C_{i,2020} + CO2\_CH_{i,2020}) + (AS_{i,t} - AO_{i,t}) \times CH_i \times CO2\_CH_{i,2020}\right). \tag{18}$$

In the following scenarios, we assume the unused rate of each city at the end of 2030 gradually reduces to 75%, 50%, and 25% of the level observed at the end of 2020. Utilizing these assumed rates of unused housing, we can impute the series of annual floor areas of housing completion using the corresponding $SUR_{i,t}$ in Eq. (17), as well as the annual series of $AS_{i,t}$. We can then calculate the total carbon emissions in the city in 2021–2030 under each scenario.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All macro-level data necessary for the calculation of unused rate and carbon emissions have been deposited in the public repository (https://gitee.com/hefanz/unused-housing-in-urban-china-and-its-carbon-emission-impact). Source data are provided with this paper.

Code availability

All codes necessary for replication have been deposited in the public repository (https://gitee.com/hefanz/unused-housing-in-urban-china-and-its-carbon-emission-impact).

References

1. Hamilton, I. et al. *2020 Global Status Report for Buildings and Construction* (United Nations Environment Programme (UNEP), 2020).
2. Li, B., Han, S., Wang, Y., Li, J. & Wang, Y. Feasibility assessment of the carbon emissions peak in China's construction industry: factor decomposition and peak forecast. *Sci. Total Environ.* **706**, 135716 (2020).
3. Mishra, A. et al. Land use change and carbon emissions of a transformation to timber cities. *Nat. Commun.* **13**, 1–12 (2022).
4. Heeren, N. et al. Environmental impact of buildings what matters? *Environ. Sci. Technol.* **49**, 9832–9841 (2015).
5. Leibowicz, B. D. et al. Optimal decarbonization pathways for urban residential building energy services. *Appl. Energy* **230**, 1311–1325 (2018).
6. You, K., Ren, H., Cai, W., Huang, R. & Li, Y. Modeling carbon emission trend in China's building sector to year 2060. *Resour. Conserv. Recycl.* **188**, 106679 (2023).
7. Berrill, P., Wilson, E. J., Reyna, J. L., Fontanini, A. D. & Hertwich, E. G. Decarbonization pathways for the residential sector in the United States. *Nat. Clim. Chang.* **12**, 712–718 (2022).
8. Zhong, X. et al. Global greenhouse gas emissions from residential and commercial building materials and mitigation strategies to 2060. *Nat. Commun.* **12**, 1–10 (2021).
9. Zhu, C., Li, X., Zhu, W. & Gong, W. Embodied carbon emissions and mitigation potential in China's building sector: an outlook to 2060. *Energy Policy* **170**, 113222 (2022).
10. Ellsworth-Krebs, K. Implications of declining household sizes and expectations of home comfort for domestic energy demand. *Nat. Energy* **5**, 20–25 (2020).
11. Wu, J., Gyourko, J. & Deng, Y. Evaluating the risk of Chinese housing markets: what we know and what we need to know. *China Econ. Rev.* **39**, 91–114 (2016).
12. Jin, X. et al. Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data. *Cities* **63**, 98–109 (2017).
13. Rosen, K. T. & Smith, L. B. The price-adjustment process for rental housing and the natural vacancy rate. *Am. Econ. Rev.* **73**, 779–786 (1983).
14. Horn, K. & Merante, M. Is home sharing driving up rents? Evidence from Airbnb in Boston. *J. Hous. Econ.* **38**, 14–24 (2017).
15. Gallent, N., Mace, A. & Tewdwr-Jones, M. *Second Homes: European Perspectives and UK Policies* (Routledge, 2017).
16. Cohen, J. R. Abandoned housing: exploring lessons from Baltimore. *Hous. Policy Debate* **12**, 415–448 (2001).
17. National Bureau of Statistics of China. *Statistical Communiqué of the People's Republic of China on the 2020 National Economic and Social Development* https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1901004.html (National Bureau of Statistics of China, 2020).
18. Rogoff, K. S. & Yang, Y. A tale of tier 3 cities. *J. Int. Econ.* **152**, 103989 (2024).
19. Saiz, A. The geographic determinants of housing supply. *Q. J. Econ.* **125**, 1253–1296 (2010).
20. Gyourko, J., Shen, Y., Wu, J. & Zhang, R. Land finance in China: analysis and review. *China Econ. Rev.* **76**, 101868 (2022).
21. Wang, X., Li, K. & Wu, J. House price index based on online listing information: the case of China. *J. Hous. Econ.* **50**, 101715 (2020).
22. Benefield, J. D., Cain, C. L. & Johnson, K. H. On the relationship between property price, time-on-market, and photo depictions in a multiple listing service. *J. Real. Estate Financ. Econ.* **43**, 401–422 (2011).
23. Luchtenberg, K. F., Seiler, M. J. & Sun, H. Listing agent signals: does a picture paint a thousand words? *J. Real. Estate Financ. Econ.* **59**, 617–648 (2019).
24. Liu, X.-Y., Wu, J. & Zhou, Z.-H. Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* **39**, 539–550 (2008).
25. Kotsiantis, S., Kanellopoulos, D. & Pintelas, P. Handling imbalanced datasets: a review. *GESTS Int. Trans. Comput. Sci. Eng.* **30**, 25–36 (2006).
26. Chawla, N. V., Japkowicz, N. & Kotcz, A. Special issue on learning from imbalanced data sets. *SIGKDD Explor. Newsl.* **6**, 1–6 (2004).
27. Jung, H. et al. ResNet-based vehicle classification and localization in traffic surveillance systems. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 934–940 (IEEE, 2017).
28. Rezende, E., Ruppert, G., Carvalho, T., Ramos, F. & de Geus, P. Malicious software classification using transfer learning of ResNet-50 deep neural network. In *Proc. IEEE International Conference on Machine Learning and Applications (ICMLA)* 1011–1014 (IEEE, 2017).
29. Habibzadeh Motlagh, M., Jannesari, M., Rezaei, Z., Totonchi, M. & Baharvand, H. Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception. In *Proc. International Conference on Machine Vision (ICMV)* (eds. Zhou, J., Radeva, P., Nikolaev, D. & Verikas, A.) 105 (SPIE, 2018).
30. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).
31. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
32. Lewis, D. D., Yang, Y., Russell-Rose, T. & Li, F. RCV1: A new benchmark collection for text categorization research. *J. Mach. Learn. Res.* **5**, 361–397 (2004).
33. Pillai, I., Fumera, G. & Roli, F. Designing multi-label classifiers that maximize F measures: state of the art. *Pattern Recognit.* **61**, 394–404 (2017).
34. Davis, J. & Goadrich, M. The relationship between Precision-Recall and ROC curves. In *Proc. International Conference on Machine Learning (ICML)* 233–240 (ACM Press, 2006).

35. Han, L. & Strange, W. C. *Handbook of Regional and Urban Economics* (eds G. Duranton, V. Henderson, & W. Strange) Vol. 5, 813–886 (North-Holland, 2015).
36. Zheng, S., Liu, H. & Lee, R. Buyer search and the role of broker in an emerging housing market: a case study of Guangzhou. *Tsinghua Sci. Technol.* **11**, 675–685 (2006).
37. Ministry of Housing and Urban-Rural Development. *Unified Standard for Reliability Design of Building Structures GB 50068-2018* (China Architecture & Building Press, 2018).
38. You, K., Yu, Y., Li, Y., Cai, W. & Shi, Q. Spatiotemporal decomposition analysis of carbon emissions on Chinese residential central heating. *Energy Build* **253**, 111485 (2021).
39. Asian Development Bank. *District cooling in the People's Republic of China: Status and Development Potential* (2017).
40. Wei, Z., Xu, J. & Wenpeng, H. The application and development of district cooling system in China: a review. *J. Build. Eng.* **50**, 104166 (2022).
41. Du, M. et al. Quantification and scenario analysis of CO_2 emissions from the central heating supply system in China from 2006 to 2025. *Appl. Energy* **225**, 869–875 (2018).
42. Zhang, Y., Xia, J., Fang, H., Jiang, Y. & Liang, Z. Field tests on the operational energy consumption of Chinese district heating systems and evaluation of typical associated problems. *Energy Build* **224**, 110269 (2020).
43. Luo, A. & Xia, J. Policy on energy consumption of district heating in northern China: Historical evidence, stages, and measures. *J. Clean. Prod.* **256**, 120265 (2020).
44. Liu, L., Fu, L., Jiang, Y. & Guo, S. Major issues and solutions in the heat-metering reform in China. *Renew. Sust. Energ. Rev.* **15**, 673–680 (2011).
45. Liu, L., Fu, L. & Jiang, Y. A new "wireless on-off control" technique for adjusting and metering household heat in district heating system. *Appl. Therm. Eng.* **36**, 202–209 (2012).
46. Yang, D. et al. Urban buildings material intensity in China from 1949 to 2015. *Resour. Conserv. Recycl.* **159**, 104824 (2020).
47. Chen, W., Yang, S., Zhang, X., Jordan, N. D. & Huang, J. Embodied energy and carbon emissions of building materials in China. *Build. Environ.* **207**, 108434 (2022).
48. Zhang, Y., Yan, D., Hu, S. & Guo, S. Modelling of energy consumption and carbon emission from the building construction sector in China, a process-based LCA approach. *Energy Policy* **134**, 110949 (2019).
49. Zou, C. et al. Toward carbon free by 2060: a decarbonization roadmap of operational residential buildings in China. *Energy* **277**, 127689 (2023).
50. Zhang, Y., He, C.-Q., Tang, B.-J. & Wei, Y.-M. China's energy consumption in the building sector: a life cycle approach. *Energy Build* **94**, 240–251 (2015).
51. Gao, T. et al. Evolution and projection of CO_2 emissions for China's cement industry from 1980 to 2020. *Renew. Sust. Energ. Rev.* **74**, 522–537 (2017).
52. Wang, K., Wang, C., Lu, X. & Chen, J. Scenario analysis on CO_2 emissions reduction potential in China's iron and steel industry. *Energy Policy* **35**, 2320–2335 (2007).
53. Chen, W., Yin, X. & Ma, D. A bottom-up analysis of China's iron and steel industrial energy consumption and CO_2 emissions. *Appl. Energy* **136**, 1174–1183 (2014).
54. Shen, J., Zhang, Q., Xu, L., Tian, S. & Wang, P. Future CO_2 emission trends and radical decarbonization path of iron and steel industry in China. *J. Clean. Prod.* **326**, 129354 (2021).
55. Teng, Y., Li, K., Pan, W. & Ng, T. Reducing building life cycle carbon emissions through prefabrication: evidence from and gaps in empirical studies. *Build. Environ.* **132**, 125–136 (2018).
56. Hao, J. et al. Carbon emission reduction in prefabrication construction during materialization stage: a BIM-based life-cycle assessment approach. *Sci. Total Environ.* **723**, 137870 (2020).
57. Yan, R., Xiang, X., Cai, W. & Ma, M. Decarbonizing residential buildings in the developing world: historical cases from China. *Sci. Total Environ.* **847**, 157679 (2022).
58. Wu, X., Peng, B. & Lin, B. A dynamic life cycle carbon emission assessment on green and non-green buildings in China. *Energy Build* **149**, 272–281 (2017).
59. Maciel, V. G. et al. Towards a non-ambiguous view of the amortization period for quantifying direct land-use change in LCA. *Int. J. Life Cycle Assess.* **27**, 1299–1315 (2022).

Acknowledgements
We highly appreciate Joseph Gyourko, Yu Qin, Xiaodong Li, and Chen Zhu for their valuable comments. Linzhen Zhu provides outstanding research assistance. All the authors are also grateful for the funding from the National Natural Science Foundation of China (Project No: 72174100, 723B2015, 72474114).

Author contributions
W.J. and Z.R. raised the research idea. Z.H. and W.J. collected data. Z.H. and Z.R. designed and trained the deep learning model. Y.X. and Z.H. calculated the carbon emissions. W.J., Z.H., Z.R., and Y.X. wrote and revised the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57217-7.

Correspondence and requests for materials should be addressed to Rongjie Zhang or Jing Wu.

Peer review information *Nature Communications* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.