nature cities

Article

https://doi.org/10.1038/s44284-024-00038-4

# The role of informal ruralization within China's rapid urbanization

Received: 24 September 2023

Accepted: 19 January 2024

Published online: 27 February 2024

Check for updates

Hanxi Wang

In China's rapidly urbanizing cities, chengshi kaihuang (CK), an informal practice of urban agriculture, has been quietly encroaching into neglected urban spaces for the cultivation of vegetables. China's unprecedented transformation from an agrarian nation to a nation of megacities over the past four decades has relied massively upon the incorporation of rural land and people in the construction and operation of its cities. Yet, while some scholars have begun to unpack the complex agency of rural bodies in China's urban environments, the rural is generally understood to be residual or obsolete against the overwhelming, top-down power of urbanization. Here we propose, through a remote sensing study of the practice, CK as an example of the bottom-up, ruralizing agency of ordinary people within China's rapid urbanization and present data on the spatial impact of CK's ruralization upon the central urban districts of Wuhan, a Chinese megacity.

In the four decades since it reopened its doors to the world, China has transformed from a nation of agriculture with a 21% urban population to a nation of megacities with a 63% urban population[1]. As the global urban population overtook the rural for the first time at the start of the twenty-first century, the UN famously declared the advent of the 'urban age'[2]. While this thesis was later criticized as inaccurate due to its problematic statistical approach and the binary categorization of the rural and the urban, succeeding works continued to reinforce the dominance of urbanization over rural territories. Most prominently in the body of scholarship inspired by Brenner and Schmid's planetary urbanization[3,4], the rural was often reduced to "an obsolescent category (a historical starting point) or residual space (an unconsumed world)"[5] under an extended planetary network of urbanization. Indeed, while planetary urbanization had illuminated the presence of urban dynamics far beyond the traditional confines of the city, this dismantling of the rural–urban dualism has produced very little evidence of rural dynamics within the newly expanded city. In 2013, Monica Krause questioned, "As we see people move into cities, why do we assume only the people to change? […] If the whole world is urbanizing, it *must* also be ruralizing"[6]. Subsequently, inspired by the growing scholarship on hybridized rural–urban phenomena in Southeast Asia[7–9], Gillen et al. argued that 'more-than-residual' processes of ruralization do exist both within what would ordinarily be considered 'urban' spaces and in relation to the processes of urbanization through the everyday practices of ordinary people who pursue ways of living that defy rural/urban binaries in the aftermath of urbanization[10].

Yet, existing scholarships on ruralization have largely focused on the fringe territories of urban development, such as the periurban or the suburban. In this Article, we present a case study that demonstrates not only the emergence of ruralization within the central districts of a rapidly growing city in China but also its potential for considerable macroscale impact upon the urban environment. Through it, this paper hopes to support and expand upon the contemporary scholarship on geographies of ruralization as proposed by Krause and Gillen et al. as well as bring attention to how ordinary people create environments beyond the institutional categorizations of rural versus urban in ways more complex than the binary of top-down versus bottom-up as they attune to the monumental transformations taking place around them.

In stagnant construction sites, neglected landscaping and forgotten spaces under viaducts and highways, China's expanding cities are secretly 'sprouting' with vegetables owing to a practice called chengshi kaihuang (CK, 城市开荒)—an informal practice where citizens infiltrate the neglected spaces of China's rapid urbanization for the cultivation of vegetables. The name CK is a combination of two terms—chengshi (城市), meaning 'urban', and kaihuang (开荒), an evocative term meaning 'the opening of wasteland.' In the long span of China's history as an agrarian nation, kaihuang has been both a pragmatic and a symbolic practice of resilience. From an ancient poet's opening of rural wasteland to escape the corruptions of the city[11] to the Communist call for self-sufficiency through wasteland cultivation during the Sino-Japanese War[12–14], kaihuang is a deceptively simple act of vegetable growing

Department of Geography, University College London, London, UK. e-mail: Hanxi.Wang.22@ucl.ac.uk

that carries within itself an almost utopic desire to survive and create alternative futures from precarity. To this day, a sculpture of the 'Kaihuang Bull' stands before the Shenzhen Municipal Party Committee Building as a symbol of the pioneering kaihuang spirit of laborers that turned a small fishing village into a megacity through hard work, perseverance and creativity[15]. In the case of CK, isolated individual citizens have turned this pioneering spirit on the wastelands of China's rapidly expanding cities and informally created thousands of acres of urban agriculture.

While Gillen et al. conceptualized ruralization as a bottom-up practice, the influence of powerful top-down actors, such as the state and global systems of labor and capital, in the spatial and discursive productions of ruralization has been a point of debate[16]. China, in particular, has been cited as an example where the geographies of ruralization are dominated by the state strategy of rural revitalization[17]. Indeed, in both rapid urbanization and ruralization, understandings of China's contemporary society have mostly been concerned with top-down actions and narratives. While some recent scholarships have begun to unpack the complex and ambiguous agency of ordinary people under China's strong state[18,19], there is a lack of empirical evidence on the spatial potential of such agency at large scales.

Existing studies of CK are few and disconnected, with individual studies focused on specific localized incidents and published under different descriptors. In two studies published subsequently in 2018 and 2020 (refs. 20,21), Zhu, He and collaborators studied the stigmatization around 'self-proclaimed vegetable lots' that have emerged in the public green spaces of residential neighborhoods in Hangzhou and suggested the potential of such bottom-up approaches as a 'new dawn of the public realm' in China, if given appropriate guidance. In 2022 (ref. 22), Asa Roast studied how the kongdi (vacant lots) of periurban Chongqing, where 'informal commoning' emerged as a 'nominally rural practice', performed as a liminal space that is simultaneously inside and outside of the city, where the act of 'becoming urban' could be negotiated.

While these studies provide valuable and detailed insights into the motivations and developmental contexts behind specific incidents of informal cultivation, they are intensive ethnographic projects focused on a limited number of physical sites. This paper leverages remote sensing and social media data to investigate the city-wide occurrence of these CK practices, as they are called by Chinese netizens, and understand their broader distribution and interaction with the urban environment of a Chinese megacity. It hopes, following the example set by scholarship such as McGee's desakota[7] and Kusno's kampungkota[23], to bring terms, ideas and creative urban practices from non-Western cultures and languages to bear on the contemporary and future understandings of rural–urban dynamics.

Over the recent decades, improvements in the quality of remote sensing data, as well as the power, efficiency and accuracy of analysis methods, have made it one of the most important tools in the study of land-use dynamics. In China, remote sensing has, among other applications, facilitated the understanding of rapid urbanization's impact across a wide range of fields, from environmental protection[24,25] to urban planning[26,27]. As far as it is possible to discover at the time of writing, this paper is the first to leverage remote sensing and social media data to study the practice of CK.

The findings of the paper, as set out below, provide a description of the spatiotemporal characteristics of CK, its presence in cities across China and, taking Wuhan as the study area, the extent to which CK has infiltrated the urban landscape of a rapidly expanding megacity.

## Results

### Spatiotemporal characteristics of CK

Unlike community gardens and allotments, CKs are illegal encroachments carried out by individuals rather than organized groups. He and Zhu observed in their first study[20] that informally claimed vegetable lots occurred in public green spaces that had 'damaged and bare surfaces' due to maintenance neglect. The 'informal commoning' in Roast's study emerged from a less ambiguous wasteland of the kongdi, tracts of demolished agricultural land that had not yet undergone development. In both instances, the sites were in neighborhoods that had undergone recent rapid urbanization.

The associations with rapid urbanization and vacant/neglected spaces are consistent with the findings from the gray literature on CK. Local news reported the occurrence of CK by railways[28], street-side landscaping[29] and, most frequently, in residential green spaces[30,31] and vacant sites nearby residential neighborhoods[32,33]. Self-narratives by practitioners on social media have mostly shown CK occurring in kongdi or huangdi (荒地, wasteland), such as neglected construction sites, storage areas and parking lots. Visually, photos and video footage of CK from both social media and news reports show the vegetable lots as patchy, irregular and small in scale (Fig. 1a). This distinct pattern makes CK easily distinguishable from other forms of urban greenery, such as parks, commercial urban agriculture and allotments, which are organized by regular grids or other geometric designs (Fig. 1, right).

News reports, social media data and interviews with participants yielded locations of CK sites in Wuhan. Evaluation of these sites through very high-resolution (VHR) satellite imagery on Google Earth Pro confirmed the physical characteristics of CK plots and revealed that CK emerged on land that has experienced the characteristic temporal sequence of rapid urbanization, whereby rural land is quickly and thoroughly demolished to bare soil (Fig. 1b and Extended Data Fig. 1).

Local news reports and government notices confirmed the description by Zhu et al.[21] that CK practices are often rejected as 'uncivil' by officials and showed that these sites are frequently demolished. VHR imagery confirms that most of the CK sites identified are demolished within 2 years of their emergence, but also shows a tendency for vegetable plots to return persistently if the demolished site remains vacant afterwards (Fig. 1c). Thus, CK can further be distinguished by its temporal instability from the continuity and formal stability of other urban green spaces.

### National occurrences and study area selection

The survey of online gray literature facilitated a preliminary assessment of the occurrence of CK at the national scale beyond isolated incidents in individual cities. Tallying city-level location information from the study's gray literature dataset, the results demonstrated that occurrences of CK are not limited to any particular region but that it is possible to find instances across the majority of provinces in China (Fig. 2). The study found fewer instances of CK in the northern provinces than central and southern provinces, where the climate is more tropical. However, it is difficult to draw any definitive conclusions without a more comprehensive survey. The gray literature search had also yielded similar practices outside of China in Vietnam, where local news have reported on informal urban gardeners who cultivate all kinds of urban edge conditions from the gaps of concrete embankments to cemeteries and public parks[34].

Within the data gathered, Wuhan was chosen as the study site for this paper as it is a major city of a reasonable size that has undergone rapid urbanization in the recent decades and has recorded a high number of CK occurrences on social media. This study is conducted over the combined area of Wuhan's central urban districts, that is, Jiang'an, Jiangshan, Qiaokou, Hanyang, Wuchang, Qingshan and Hongshan. The capital city of Hubei Province, Wuhan (29° 58′ to 31° 22′ N, 113° 41′ to 115° 05′ E), is located at the intersection of the Yangtze River and the Han River. The two rivers meeting at the center of the city and the presence of hundreds of lakes around them have earned Wuhan the name of 'the city of a hundred lakes.' As the most populous city in central China, Wuhan saw, from 2000 to 2019, a massive increase in built-up area of 982.66 km$^2$ (228% of its original size), along with a reduction of 717.14 km$^2$ of cropland[26]. Wuhan's economic boom occurred in the

https://doi.org/10.1038/s44284-024-00038-4

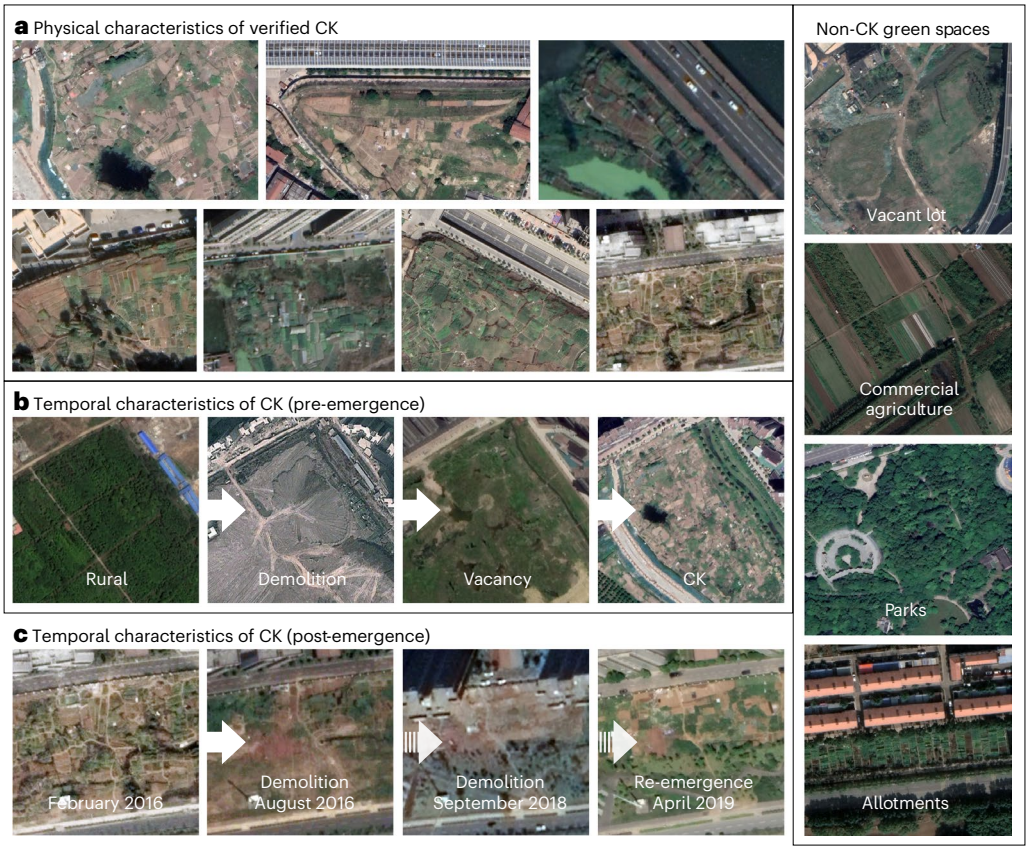



**Fig. 1 | Identifying characteristics of CK in remote sensing data. a**–**c**, Left: examples showing typical physical characteristics of CK, that is, small, patchy, irregular agricultural plots in urban contexts (**a**); an example of the typical land-use sequence that precedes the emergence of CK (**b**); and an example of the land-use pattern after CK emergence (**c**). Right, examples of how similar but non-CK urban land use types can be visually distinguished from CK. Map data retrieved from Google, 2023 Maxar Technologies.

twenty-first century as the site of multiple urban innovation initiatives. It was declared the first pilot area of the 'two-oriented society'[35], which targets economic development along with resource conservation and environmental protection. It has also been the site of multiple state-designated technology development zones such as the Donghu New Technology Development Zone (also called the Optics Valley, http://www.chinaopticsvalley.com/) to the east and the Wuhan Economic and Technological Development Zone (http://en.whkfq.gov.cn/) to the south.

## CK in Wuhan, 2017–2022

Owing to the lack of existing information on CK sites, the study employed a 'snowballing' method to conduct a city-wide survey by multispectral remote sensing data. Analysis of the gray literature yielded seven suitable CK sites for the initial supervised machine-learning land cover classification (LCC). For each year between 2017 and 2022, an initial LCC was trained to obtain a preliminary mapping of potential CK sites. Sampling the initial LCC mapping then yielded 150–300 additional verified CK sites per year, with which it was possible to conduct a much more accurate final LCC of the study area.

In general, the initial LCCs detected less CK sites (averaging 4.09% of the study area) than the final LCCs. Though the initial LCC had limited access to training points, it yielded relatively high values in its accuracy measures, which may have been skewed owing to the equally limited number of validation points available to produce the error matrix (Extended Data Table 5). The final LCCs (Fig. 3), which were trained with 400–800 CK points and 11,000–17,000 non-CK points depending on the year, yielded higher numbers for the overall area of CK sites in the study area as well as slightly higher accuracy measures.

The final LCC showed that the average percentage of Wuhan's central urban districts with active CK between 2017 and 2022 is 6.34%, constituting 12,941 acres of land. The study area contained the lowest amount of CK activity in 2018 and the highest amount in 2021 (Table 1).

Similar to, and potentially even more so than other forms of urban agriculture, CK sites possesses characteristics such as high variation in phenology, small and irregular sites and fragmentation due to the dense heterogeneity of urban environments[36,37], that make it difficult for the classifier to distinguish them from other urban land uses. Furthermore, the study result is limited by the 10 m resolution of Sentinel-2 multispectral data, which may at times be too coarse for the detection of small or narrow CK sites.

For most years, the final LCC achieved an overall accuracy above 98% and a kappa index above 0.80. Both the initial and final LCCs detected non-CK sites with more accuracy than CK sites. The error matrix showed higher consumer accuracies than producer accuracies for CK sites, which indicates that the classifier was more successful at identifying CK sites accurately than comprehensively. For the final LCCs, the lowest producer accuracy for CK sites (62.2%) was recorded in 2018, which detected the smallest total of CK area between 2017 and 2022. Conversely, the lowest consumer accuracy for CK sites was recorded in 2020, which detected the largest total of CK area between 2017 and 2022 (Extended Data Table 5).

## Emergence types and interactions with the city

Four 3 km by 3 km sample areas of the LCC mapping were randomly selected for detailed visual assessment using VHR satellite imagery. The sampling yielded 544 CK sites whose land-use histories both before and after emergence were evaluated. The analysis of land-use history

https://doi.org/10.1038/s44284-024-00038-4

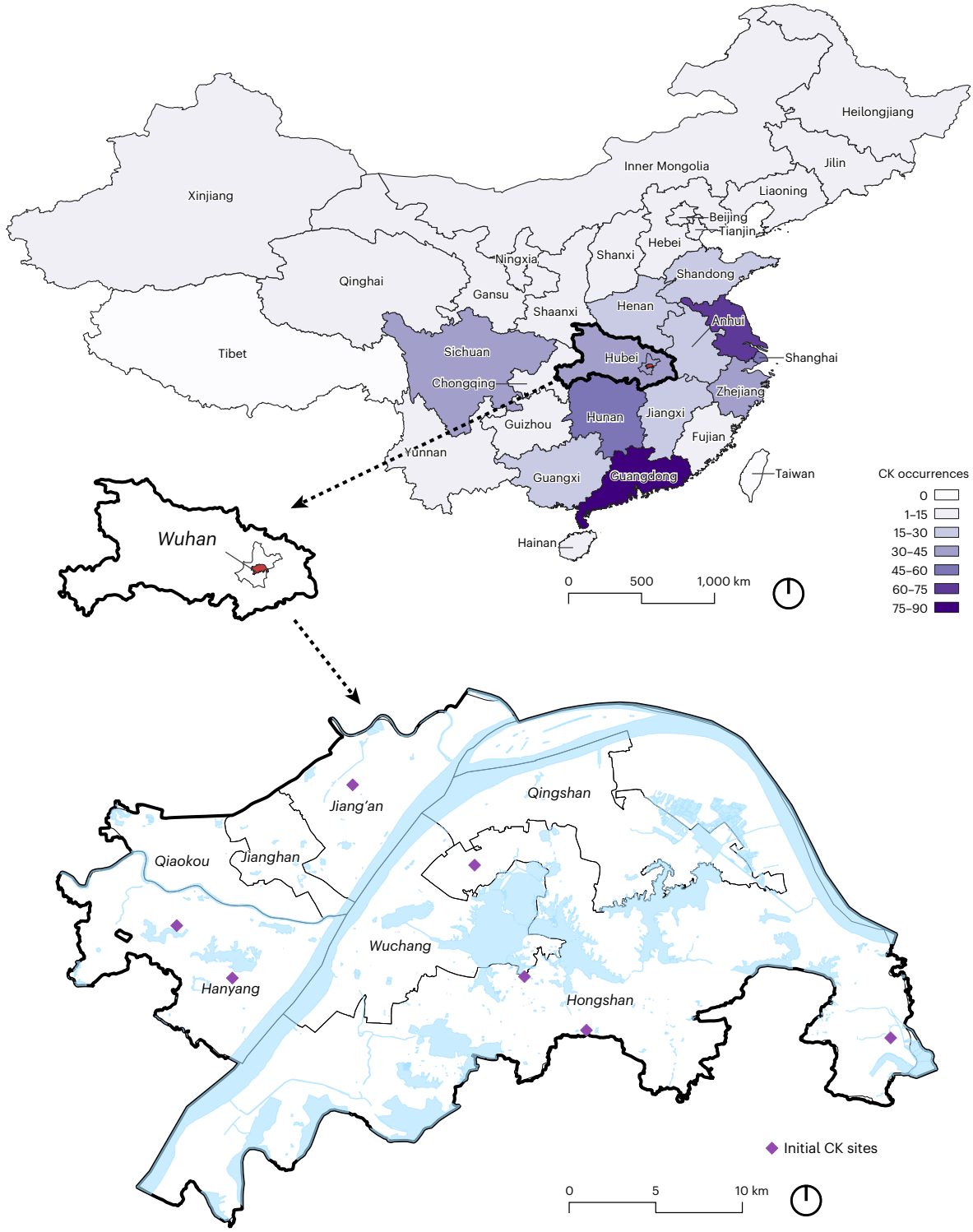



**Fig. 2 | Assessment of CK occurrences from gray literature.** Top, distribution of CK sites found in China and the study area. Bottom, the metropolitan regions of Wuhan with locations of initial CK sites.

before emergence revealed five types of urban conditions where CK has occurred.

A total of 97.4% of the sites had undergone typical processes of rapid urbanization post-2000 (Extended Data Fig. 1), and 66.7% were land that had been vacant since its demolition from rural land, 19.7% were undeveloped land that were used as construction staging areas at some point and 11.0% were the irregular parcels left by the construction of transportation infrastructure. The remaining 2.6% of the sites consisted of 1.8% in existing urban edge conditions (for example, riverbanks) and 0.7% in existing urban green spaces (for example, parks; Figs. 4a and 5a).

Studies have shown that rapid urbanization is one of the leading causes of urban vacant lands in developing cities owing to factors such as capital shortage, land speculation and urban planning adjustments[38,39]. The satellite data of CK sites showed demolition of agricultural farmland, fishponds and residential clusters on a massive scale for

https://doi.org/10.1038/s44284-024-00038-4

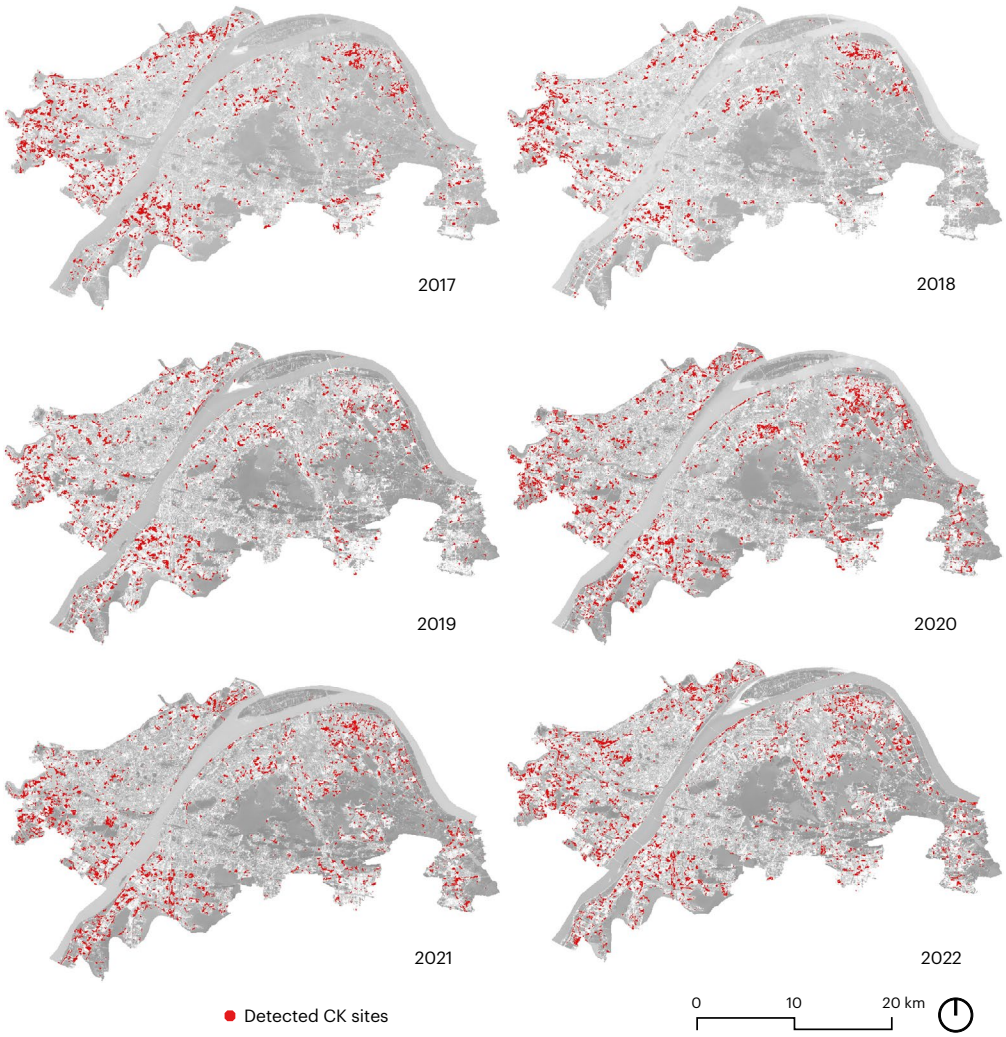



**Fig. 3 | Final supervised LCC results, 2017–2022.** Each mapping, denoted by year, shows, in red, the areas detected as CK within the metropolitan districts of Wuhan.

**Table 1 | Results and accuracy measures of the initial and final LCCs, 2017–2022**

| Year | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|
| Initial LCC results | | | | | | |
| CK area (acres) | 7,008 | 6,120 | 12,192 | 9,903 | 10,006 | 4,930 |
| Percentage of city area | 3.43% | 3.00% | 5.97% | 4.85% | 4.90% | 2.41% |
| Kappa index | 0.81 | 0.74 | 0.72 | 0.78 | 0.89 | 0.79 |
| Overall accuracy | 96.76% | 94.64% | 97.57% | 96.70% | 98.33% | 96.72% |
| Final LCC results | | | | | | |
| CK area (acres) | 13,968 | 8,540 | 10,489 | 15,730 | 14,071 | 14,848 |
| Percentage of city area | 6.84% | 4.18% | 5.13% | 7.70% | 6.89% | 7.27% |
| Kappa index | 0.83 | 0.75 | 0.83 | 0.84 | 0.80 | 0.86 |
| Overall accuracy | 98.57% | 98.83% | 98.97% | 98.45% | 97.75% | 98.87% |

urban development. The expansion of the city also saw the encroachment of roads, highways, tunnels and elevated railways directly into rural lands with little regard for existing contexts, which disrupted agricultural production and left behind isolated and irregular edge lands that are difficult to develop but perfect for the infiltration of CK practitioners. Only 8% of CK emerged on sites left vacant for less than 2 years, while 66% of the sites had been left vacant for between 2 and 5 years before practitioners moved in and 26% had been left vacant for 6 years or longer.

The examination of land use after the emergence of CK showed that 24.4% have been redeveloped as buildings and 5.9% as parks or other urban green spaces, while 69.7% of the sites remain undeveloped

https://doi.org/10.1038/s44284-024-00038-4

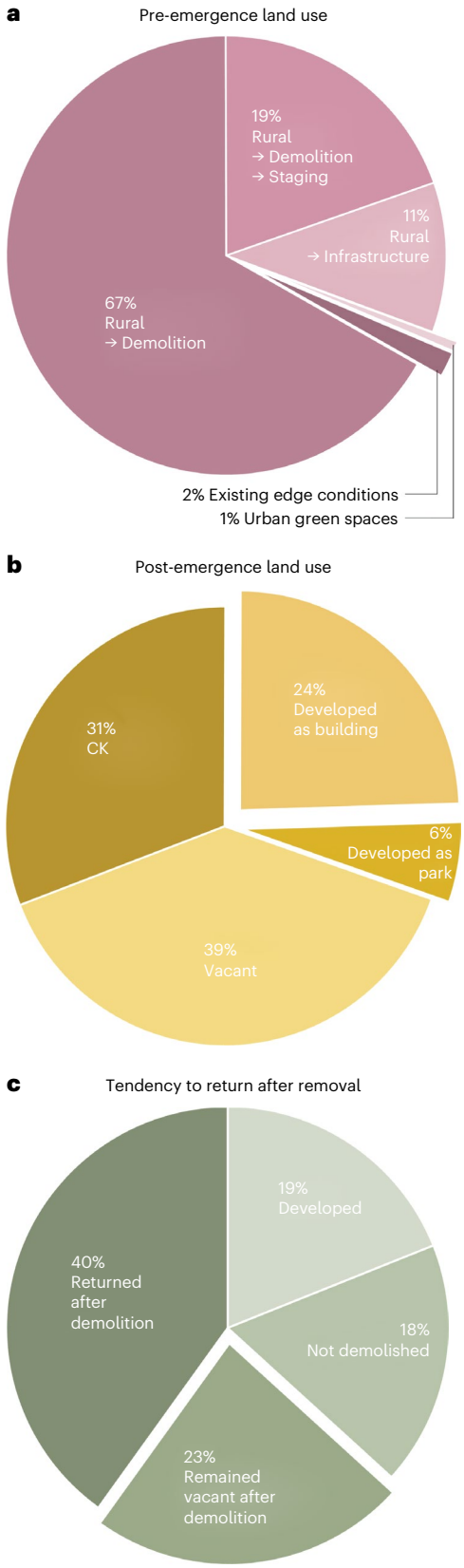



**Fig. 4 | Distributions of land-use characteristics within sampled sites. a**, Land-use histories before CK emergence. **b**, Land-use histories after CK emergence. **c**, Tendency for CK to return after forced removal.

(Figs. 4b and 5b). The study also found that 82.2% of CK sites experience forced removal and demolition, which includes 19% where the site was immediately developed, 23% where the site remained vacant but CK never returned and 40% where CK continued to return until the land was developed (Figs. 4c and 5c).

### Social media representations and discourses

Beyond the physical conditions, the contextual analysis of gray literature indicates that the rapidity of China's urbanization may also be connected to psychological motivations for practitioners of CK. Analysis of the social media posts collected in this study showed that the city is most often described as 'loud/continuous noise' ('喧嚣'), 'every inch is expensive' ('寸土寸金'), 'foreign' ('陌生') and a place of 'struggle' ('拼搏'). CK is presented as a rare escape of 'country lifestyle' ('田园生活') that offers peace and relaxation from such urban harshness. Posters described CK as a practice that offered a wide array of rare benefits for an urban dweller—the simple pleasure of working with one's own hands, proximity to nature, access to healthy organic vegetables, outdoor exercise, a sense of freedom and autonomy and engagement with rural childhoods and pasts, as well as the ability to save money by becoming self-sufficient in vegetables.

The posts frequently incorporated popular political and cultural slogans to allude to moral values such as industriousness—'The Laborer is the Most Beautiful/Honorable' ('劳动人民最光荣'/ '劳动者是最美的人') and 'A Bit of Cultivation Is a Bit of Harvest' ('一份耕耘一份收获'); self-sufficiency—'Do It Yourself, Want for Nothing' ('自己动手丰衣足食') and 'Work Hard with Your Own Hands for Prosperity' ('勤劳致富靠双手'); ecological civilization—'My Home Is Green Waters and Lush Mountains' ('绿水青山我的家'); and rural revitalization—'2022 New Farmer Strategy' ('2022新农人计划') and 'Three Rural Problems' ('三农'), covering ideologies that speak to a wide range of topics from both the past and present of Chinese society and the Chinese Communist Party.

## Discussion

The LCC results estimated that an average of 12,941 acres of CK were present in Wuhan's metropolitan regions (204,260 acres, 826.6 km$^2$) between 2017 and 2022. This figure far exceeds the known scope of urban agriculture in most comparably sized cities—approximately 120 acres of community gardens within the five boroughs of New York City (193,681 acres, 783.8 km$^2$)[40,41], 65 acres (264,181m$^2$) of both private and public urban agriculture in the city of Chicago (149,746 acres, 606 km$^2$)[42] and 1,606 acres of backyard and community gardens in Toronto (155,726 acres, 630 km$^2$)[43]—but falls below the estimated 87,000 acres in Havana (179,942 acres, 728.2 km$^2$)[44]. While the informal nature and lack of organization had distinguished CK from common types of urban agriculture such as community gardens, allotments, urban farms and home gardens, this study has shown that CK possesses an astonishing potential to create large-scale spatial transformations that greatly exceeds current understandings of informal urban agriculture such as guerrilla gardening[45,46].

The land-use histories of detected CK sites seen through time-series VHR imagery demonstrate how almost all have emerged from sites of large-scale demolition for urban expansion. They also show that after forced removal, CK is twice as likely to return than not if the site remains undeveloped. Thus, this paper argues that CK is responding to the specific spatial conditions created by the inefficiencies of rapid urbanization. Zhu et al. recognized self-claimed vegetable lots as a way for Hangzhou citizens to claim their 'right to the city' without waiting for an invitation to engage in governance[21]. The wider occurrence of CK detected through gray literature suggests that similar rights are potentially being claimed at the national scale by individuals who are acting independently but concurrently.

Discursively, the contextual analysis showed that, while state officials and most local news outlets stigmatize CK as 'unmannered and uncivilized'[21], the practitioners' social media posts have leveraged a wide range of both past and present, popular and state-sponsored narratives to represent their encroachment as culturally, if not legally, valid. While rural support played a crucial role in the victory of the Communist party during the civil war and was widely valorized in Maoist

https://doi.org/10.1038/s44284-024-00038-4


**Fig. 5 | Examples illustrating land-use characteristics of sampled CK sites before and after the emergence of CK. a**, Examples of the five pre-emergence land-use sequences: left to right, rural to demolition, rural to construction staging, rural to infrastructure, urban green spaces, edge conditions. **b**, Examples of post-emergence land-use sequences: left, demolition and development to building, right, demolition and development to trees/park. **c**, Example of post-emergence land-use sequence where CK is demolished but returns repeatedly. Map data retrieved from Google©, 2023 Maxar Technologies.

ideology for national movements such as the Cultural Revolution, modern China's developmental strategy relied upon a dualist rural/urban governance that privileged urban growth at the cost of rural resources[47]. Since then, the rural became commonly represented as 'poor', 'backward' and 'uncivil'[48]. Since 2017, China's rural revitalization strategy has sought to reframe the country's urban–rural relationship, issuing calls such as 'understand agriculture, love the peasantry and love the countryside' in hopes of attracting urban investments and migration back to the villages[17].

References to the imaginaries revived in the national strategy for rural revitalization frame the practice of CK as part of the 'beautiful countryside' and the rural root of Chinese identity. While hashtags such as 'My Home Is Green Waters and Lush Mountains' and 'The Laborer Is the Most Beautiful/Honorable' draw directly from other contemporary policies such as ecological civilization[49] and national self-sufficiency[50], CK practitioners also draw extensively from political movements that had valorized the rural in the past. Like the practitioners' chosen nomenclature of 'kaihuang', references to the Maoist slogans of agrarian socialism such as 'Do It Yourself, Want for Nothing' evokes a nostalgic sense of national identity and civic duty that has remained important in contemporary Chinese society.

Thus, the sites of the ruralization of CK serve not only as spaces where the inside and outside of the urban are negotiated[22] but also where the past and present visions of China's future can be arbitrated to benefit individual aspirations for better living. This paper has demonstrated the ruralizing impact of CK, an informal practice of isolated individuals, upon the urban center of the megacity of Wuhan. Given its scalar potential, further studies on the ecological, social and economic impact, as well as the spatial distribution of the practice beyond this paper's study area of Wuhan, are likely to yield interesting findings on a range of important contemporary issues from urban metabolism and sustainability to food insecurity and spatial inequality.

Though CK is correlated to specific characteristics of China's urban growth, the issues of rapid urbanization it speaks to are widespread across the developing world, as are informal urban practices of self-help and agriculture that have often been dismissed as signs of incomplete development[45]. Beyond the emergent ruralization within China's rapid urbanization, the findings of this study argue also for the importance of examining urban practices beyond the binaries of top-down versus bottom-up, urban versus rural and subject versus resistance, as they often obscure the subtle nuances of agency and their potential to contribute to inclusive, resilient and sustainable designs of the built environments at the regional, and potentially global, scale.

## Methods

### Social media data collection and contextual analysis

For contextual understanding of CK, this paper drew from existing studies and a compiled dataset of public news reports, local government

notices, social media posts and online interviews of practitioners gathered as part of a postgraduate research dissertation at the School of Geography and the Environment of the University of Oxford entitled 'Ruralizing Urbanization: Homesteads, Subjectivities and Subversions in China's Growing Cities' (ethics approval reference SOGE1A2021-110). News reports and government notices were collected by searching for the phrase '城市开荒种菜' (CK for vegetables) on Baidu News, a popular Chinese news search engine. Social media posts were collected by searching the same phrase on Douyin, the Chinese equivalent of TikTok and one of the most popular social media platforms in the country. The data collected were analyzed by manually extracting any descriptions or content related to the three areas of interest below.

The first is any information relating to spatial, temporal or physical characteristics of CK that can be used to identify it through remote sensing data. This may be verbal or text-based descriptions of land-use patterns of CK sites before cultivation as well as images or videos that show what CK sites and their surroundings look like physically. The descriptions, images and videos extracted could be sorted into common patterns, which were summarized to provide an understanding of the spatial and temporal characteristics of CK (see 'Results') that informed the procedure for the visual verification of potential sites via satellite imagery as well as the design of the remote sensing dataset for classification.

The second area of interest is the geographic locations of CK sites. While news reports and government notices always provide the city, and often the neighborhood or street address, of the site, Douyin users can choose whether to tag the city that they are in on their posts. Some users disclose more detailed information verbally within their posts or in replies to comments. City- and province-level locations were tallied in the national distribution map in Fig. 1. Sites found in Wuhan were assessed via VHR satellite imagery to verify their existence and suitability as training data for the land cover classifier (see description of this assessment in 'Visual verification of CK sites through VHR imagery' section).

Finally, the third area of interest covers any social, political or cultural references associated with CK and the city in gray literature. The researcher cross-checked any hashtags, slogans and repeated phrases within post captions and descriptions with academic databases, as well as Chinese gray literature, to identify whcih references are used by the content creators, how their use aligns with/differs from conventional uses and how this use frames the practice of CK.

### Visual verification of CK sites through VHR imagery

Manual visual interpretation of remote sensing imagery has been a fundamental part of the field[42,51]. Multiple studies have used this method on VHR imagery to obtain training data for LCC[52,53]. The Google Earth Pro platform contains a repository of historical VHR imagery from both satellite data and aerial photogrammetry that facilitated the visual verification of the temporal and spatial characteristics of CK. For the study area, VHR images going back as far as 2000 (Extended Data Table 1) were used for analysis. Each potential site was assessed first by its physical appearance in the year where the CK is active, then by its land-use history (Fig. 2). An active CK resembles urban agriculture in its vegetation, but where the spatial pattern of allotments and commercial farms are usually regular and geometrical, CK is random, composed of small plots of varying shapes and orientation. Temporally, the land use preceding CK should show a period of vacancy (usually after demolition) before the establishment of small plots of agriculture that emerge gradually in a rhizomic manner. A more detailed description of the distinguishing characteristics of CK can be found in Results.

### Supervised machine-learning LCC

Remote sensing data have been recognized as one of the most powerful tools available in the monitoring of global land cover changes[54,55]. In the case of China's rapid urbanization, remote sensing methods have made it possible to understand its diverse and multiscalar impacts, such as the assessment of environmental changes[56–58], urban land-use dynamics[26,53,59], agricultural productivity[60], etc.

Additionally, the bird's-eye view facilitated by remote sensing methods and the frequency of its data collection have proved useful in the understanding of informal phenomena, such as informal settlements, which are usually difficult to study owing to the lack of official documentation and the intensive labor associated with in-person surveys[61–63]. In recent years, the increasing availability of high-resolution satellite data has also made it possible to apply remote sensing to the detection and monitoring of small land-use phenomena such as urban agriculture[36,42,64,65]. In the context of this study, the remote sensing method of supervised machine-learning LCC facilitates a macrolevel understanding of CK that would not have been possible through conventional ethnographic methods (see Extended Data Fig. 2 for the LCC design).

As CK is an informal practice with little existing research or documentation, this study had access to only a small pool of training data for the LCC. A study by Ramezan et al. of the effect of training set size on supervised machine-learning LCC showed that larger training sizes generally yielded more accurate results but also demonstrated that the random forest and gradient-boosted trees algorithms could yield reasonably high accuracies even with a very small training size[66]. As such, this study used the random forest algorithm, which has generally proven to be one of the most efficient and accurate algorithms for LCC[66,67].

As this study is only concerned with the detection of CK sites, just two land cover classes are included: CK and non-CK sites (Extended Data Table 2). Seven CK sites (yielded from the gray literature, social media data and practitioner interviews) and a collection of non-CK sites (a selection of existing land uses such as trees, commercial farms, buildings, water bodies, etc. obtained from OpenStreetMap) provided between 770 and 1,100 data points, which are randomly allocated 70% for training and 30% for validation (Extended Data Table 5).

Selective manual visual interpretation of the LCC results mapped onto Google Earth Pro's VHR satellite imagery yielded 544 additional verified CK sites within four 3 m by 3 m sample areas, each located in the north, east, south and west sections of the city. To further overcome the initial lack of sample sites, these additional sites were added to create an enhanced training dataset that was used to train a second LCC with between 11,500 and 17,880 data points. This process allowed the city-wide survey of CK with good accuracy despite a small initial sample size (Table 1 and Extended Data Table 5).

### Remote sensing data acquisition and preprocessing

This study leveraged high-resolution multispectral satellite images acquired by the European Space Agency's Sentinel-2 mission[68] between April 2017 and October 2022. The Sentinel-2 data were accessed through the Google Earth Engine, a popular cloud-based computing platform where users can efficiently analyze large quantities of remote sensing data through a web-based code editor[55,69]. For each year from 2017 to 2022, all Sentinel-2 data covering the study area with less than 5% cloud cover are collected to form a cloud-masked median composite image. Seven 10 m and 20 m spectral bands are extracted from the composite images to form the basis of the training data for the LCC.

Seven spectral indices that provide information related to CK are added to each composite image, including vegetation indices (Normalized Difference Vegetation Index[70], Enhanced Vegetation Index[71]), the Bare Soil Index[72], the Normalized Difference Water Index[73] and the Normalized Difference Latent Heat Index[74], as well as a multitemporal band measuring the Bare Soil Index standard deviation from 2001 to the year of the composite image generated using Landsat 5 and 8 data (Extended Data Tables 3 and 4).

This study uses an object-oriented approach, which has been demonstrated to improve the accuracy of the trained classifiers for

urban land uses and smallholder agriculture[36,75]. Two object-oriented methods are applied to the remote sensing data. The first is the segmentation of composite images using object-based image analysis[76,77]. Object-based image analysis uses a segmentation algorithm to group similar pixels into object-based clusters, which makes it possible to assess object properties such as shapes and is helpful in preventing noise during classification. The second is the calculation of textural measures, which help the classifier recognize differences in spatial patterns[78]. For this study, nine textural features were calculated using the gray-level co-occurrence matrix[79,80] and incorporated into one band using principal component analysis[81].

The resulting dataset is a segmented composite multispectral image for each year between 2017 and 2022, each containing 16 bands of spectral and textural information, ready for LCC.

### Classifier accuracy assessment

For both the initial and the second LCC conducted each year between 2017 and 2022, 70% of the sample data points (for both CK and non-CK sites) are randomly chosen for the training dataset, and the remaining 30% are used for validation. Each LCC result was assessed against the validation dataset and reported in a confusion matrix. Four accuracy measures were calculated from the confusion matrix—the producer accuracy and consumer accuracy for each land cover (Extended Data Table 5), the overall accuracy and Cohen's kappa coefficient for the overall classifier (Table 1).

### Reporting summary

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

## Data availability

This study leverages publicly available satellite data from the European Space Agency and Google Earth Pro. Other data collected and generated in this study have been uploaded to https://figshare.com/projects/Cities_Gone_Wild_Understanding_Informal_Ruralization_Within_China_s_Rapid_Urbanization_2023-12-14_/189495. This includes the rasterized results of the supervised LCC for CK sites, a spreadsheet of the land-use histories of sampled CK sites and a spreadsheet listing national occurrences of CK collected from gray literature (social media posts by nonpublic entities are not included to protect individual anonymity).

## Code availability

The code used to conduct the supervised machine-learning LCC on Google Earth Engine is available at https://code.earthengine.google.com/?accept_repo=users/hw422/Wang_2024.

## References


1. Xu, X. & Akita, N. Demolition/reconstruction, and comprehensive renovation? Reflections on the renewal of urban villages in North China: a case study of a Beijing urban village. *Int. Rev. Spat. Plan. Sustain. Dev.* **9**, 62–75 (2021).
2. United Nations Population Fund. *State of World Population 2007: Unleashing the Potential of Urban Growth* (United Nations, 2007); https://doi.org/10.18356/fe74b223-en
3. Brenner, N. *Implosions/Explosions: Towards a Study of Planetary Urbanization* (Jovis, 2014).
4. Ruddick, S., Peake, L., Tanyildiz, G. S. & Patrick, D. Planetary urbanization: an urban theory for our time? *Environ. Plan. Soc. Space* **36**, 387–404 (2018).
5. Gillen, J., Bunnell, T. & Rigg, J. Geographies of ruralization. *Dialogues Hum. Geogr.* **12**, 186–203 (2022).
6. Krause, M. The ruralization of the world. *Public Cult.* **25**, 233–248 (2013).
7. McGee, T. G. in *International Encyclopedia of Geography* 1–2 (Wiley, 2017).
8. Bunnell, T. Kampung rules: landscape and the contested government of urban(e) Malayness. *Urban Studies* **39**, 1685–1701 (2002).
9. McGrath, B., Sangawongse, S., Thaikatoo, D. & Corte, M. B. The architecture of the metacity: land use change, patch dynamics and urban form in Chiang Mai, Thailand. *Urban Plan.* **2**, 53–72 (2017).
10. Gillen, J., Bunnell, T. & Rigg, J. Beyond binaries? Spatial possibilities in Southeast Asia. *Dialogues Hum. Geogr.* **12**, 227–231 (2022).
11. Kim, H. P. *Oh, Let Me Return!: Nature's Poets–Chinese Poetry of Two Millennia* (Liverpool University Press, 2017).
12. Muscolino, M. S. Refugees, land reclamation, and militarized landscapes in wartime China: Huanglongshan, Shaanxi, 1937–45. *J. Asian Stud.* **69**, 453–478 (2010).
13. Lawson, J. Unsettled lands: labour and land cultivation in western China during the War of Resistance (1937–1945). *Mod. Asian Stud.* **49**, 1442–1484 (2015).
14. Shi, X. Explore the practical path forward the spirit of Nanniwan farming for self-reliance. In *Proc. SHS Web of Conferences* Vol. 163 (EDP Sciences, 2023).
15. Xu, Y. *How did the Kaihuang Bull Become Shenzhen's Symbol? Art Classics will Tell You* [in Chinese] (CCTV News, 15 March 2021); https://news.cctv.com/2021/03/15/ARTIDMjsbWmXKeqM9BcVM74c210315.shtml
16. Ghosh, S. In what sense ruralization? *Dialogues Hum. Geogr.* **12**, 208–212 (2022).
17. Chen, N. & Kong, L. Rural revitalization in China: towards inclusive geographies of ruralization. *Dialogues Hum. Geogr.* **12**, 213–217 (2022).
18. Lora-Wainwright, A. Grassroots perspectives on relocation: threats and opportunities. *Positions Asia Crit.* **22**, 661–689 (2014).
19. Richaud, L. & Amin, A. Life amidst rubble: migrant mental health and the management of subjectivity in urban China. *Public Culture* **32**, 77–106 (2020).
20. He, B. & Zhu, J. Constructing community gardens? Residents' attitude and behaviour towards edible landscapes in emerging urban communities of China. *Urban For. Urban Green* **34**, 154–165 (2018).
21. Zhu, J., He, B.-J., Tang, W. & Thompson, S. Community blemish or new dawn for the public realm? Governance challenges for self-claimed gardens in urban China. *Cities* **102**, 102750 (2020).
22. Roast, A. Theory from empty land: informal commoning outside/within economies and ecologies of the urban. *Int. J. Urban Reg. Res.* **46**, 387–404 (2022).
23. Kusno, A. Middling urbanism: the megacity and the kampung. *Urban Geogr.* **41**, 954–970 (2020).
24. Yang, T. et al. Monitoring ecological conditions by remote sensing and social media data—Sanya City (China) as case study. *Remote Sens.* **14**, 2824 (2022).
25. Jin, Y., Liu, X., Chen, Y. & Liang, X. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. *Int. J. Remote Sens.* **39**, 8703–8723 (2018).
26. Zhai, H. et al. Understanding spatio-temporal patterns of land use/land cover change under urbanization in Wuhan, China, 2000–2019. *Remote Sens.* **13**, 3331 (2021).
27. Lang, W., Deng, J. & Li, X. Identification of 'growth' and 'shrinkage' pattern and planning strategies for shrinking cities based on a spatial perspective of the pearl river delta. *Region. J. Urban Plan. Dev.* **146**, 05020020 (2020).
28. Zhang, M. *'Urban Farmers' Kaihuang for Vegetables by Railway Tracks Safety Protection Zone Became 'Happy Farms'* [in Chinese] (Xi'an News, 18 February 2017); https://www.cnr.cn/sxpd/sx/20170218/t20170218_523608110.shtml

29. Liunan District Urban Management and Law Enforcement. *Casual Kaihuang for Vegetables is Unacceptable; City Management Removes Many Sites of 'Vegetable Gardens* [in Chinese] (Liuzhou News, 21 March 2023); http://www.liunan.gov.cn/xwzx/lndt/t19700101_3240735.shtml
30. Zhang, H. *Nanjing Qixia Maqun City Establish 'Kaihuang' Special Management Around Communities* [in Chinese] (Yangzi Evening News, 16 December 2021); https://baijiahao.baidu.com/s?id=1719300385217737440&wfr=spider&for=pc
31. *Urban Green Spaces 'Kaihuang' for Vegetables Legally Removed by Hefei Binhu Shiji Community Management* [in Chinese] (Anhui General Fenghuang Web, 7 May 2018); http://ah.ifeng.com/a/20180507/6556117_0.shtml
32. Lu, Y., Wang, R., Xiang, Q. & Zou, B. *The Neighbourhood has been Kaihuang-ed Again? Authority's Gentle Management Solves Problem* [in Chinese] (Guiyang Daily, 7 August 2023); https://baijiahao.baidu.com/s?id=1773571787759665238&wfr=spider&for=pc
33. Dai, Q. *Residential Development Vacant for 2–3 Years Residents Kaihuang to Create 'Farms'* [in Chinese] (China News, 15 December 2011); https://www.chinanews.com.cn/estate/2011/12-15/3533894.shtml
34. *Many Unique Styles of Growing Clean Vegetables of Hanoi People* [in Vietnamese] (Dân Trí, 18 Jun 2016); https://dantri.com.vn/xa-hoi/muon-kieu-trong-rau-sach-doc-dao-cua-nguoi-ha-noi-20160617031204069.htm
35. Chen, X., Liu, X. & Hu, D. Assessment of sustainable development: a case study of Wuhan as a pilot city in China. *Ecol. Indic.* **50**, 206–214 (2015).
36. Dupuy, S. et al. Analyzing urban agriculture's contribution to a southern city's resilience through land cover mapping: the case of Antananarivo, capital of Madagascar. *Remote Sens.* **12**, 1962 (2020).
37. Mansaray, L. R., Huang, W., Zhang, D., Huang, J. & Li, J. Mapping rice fields in urban Shanghai, Southeast China, using Sentinel-1A and Landsat 8 datasets. *Remote Sens.* **9**, 257 (2017).
38. Zhang, W., Li, Y. & Zheng, C. The distribution characteristics and driving mechanism of vacant land in Chengdu, China: a perspective of urban shrinkage and expansion. *Land Use Policy* **132**, 106812 (2023).
39. Qu, Y., Jiang, G., Li, Z., Shang, R. & Zhou, D. Understanding the multidimensional morphological characteristics of urban idle land: stage, subject, and spatial heterogeneity. *Cities* **97**, 102492 (2020).
40. *Community Gardens Database* (GrowNYC, 2015).
41. Gittleman, M., Farmer, C. J. Q., Kremer, P. & McPhearson, T. Estimating stormwater runoff for community gardens in New York City. *Urban Ecosyst.* **20**, 129–139 (2017).
42. Taylor, J. R. & Lovell, S. T. Mapping public and private spaces of urban agriculture in Chicago through the analysis of high-resolution aerial images in Google Earth. *Landsc. Urban Plan.* **108**, 57–70 (2012).
43. MacRae, R. et al. Could Toronto provide 10% of its fresh vegetable requirements from within its own boundaries? Matching consumption requirements with growing spaces. *J. Agric. Food Syst. Community Dev.* **1**, 105–127 (2010).
44. Koont, S. The urban agriculture of Havana. *Mon. Rev.* https://doi.org/10.14452/MR-060-08-2009-01_5 (2009).
45. WinklerPrins, A. M. G. A. *Global Urban Agriculture: Convergence of Theory and Practice Between North and South* (CABI, 2017).
46. Bach, C. E. & McClintock, N. Reclaiming the city one plot at a time? DIY garden projects, radical democracy, and the politics of spatial appropriation. *Environ. Plan. C Polit. Space* **39**, 859–878 (2021).
47. Wong, S. W., Tang, B. & Liu, J. Rethinking China's rural revitalization from a historical perspective. *J. Urban Hist.* **48**, 565–577 (2022).
48. Lai, L. *Hygiene, Sociality, and Culture in Contemporary Rural China* (Amsterdam Univ. Press, 2016).
49. Krischer, O. & Tomba, L. *Shades of Green—Notes on China's Eco-civilisation* (Eds Krischer, K. & Tomba, L.) (Made in China Journal, 2019).
50. Davidson, H. *Xi Jinping Urges China to Greater Self-reliance Amid Sanctions and Trade Tensions* (The Guardian, 2023); https://www.theguardian.com/world/2023/mar/06/xi-jinping-urges-china-to-greater-self-reliance-amid-sanctions-and-trade-tensions
51. Lillesand, T. M., Kiefer, R. W. & Chipman, J. W. *Remote Sensing and Image Interpretation* (Wiley, 2015).
52. Macarringue, L. S. et al. Land use and land cover classification in the northern region of Mozambique based on landsat time series and machine learning. *ISPRS Int. J. Geo-Inf.* **12**, 342 (2023).
53. Cai, G. et al. Detailed urban land use land cover classification at the metropolitan scale using a three-layer classification scheme. *Sensors* **19**, 3120 (2019).
54. Li, C. et al. The first all-season sample set for mapping global land cover with Landsat-8 data. *Sci. Bull.* **62**, 508–515 (2017).
55. Phan, T. N., Kuch, V. & Lehnert, L. W. Land cover classification using Google Earth Engine and Random Forest Classifier—the role of image composition. *Remote Sens.* **12**, 2411 (2020).
56. Xiong, Y. et al. The impacts of rapid urbanization on the thermal environment: a remote sensing study of Guangzhou, South China. *Remote Sens.* **4**, 2033–2056 (2012).
57. Chen, Y. et al. Spatial heterogeneity of vegetation phenology caused by urbanization in China based on remote sensing. *Ecol. Indic.* **153**, 110448 (2023).
58. Xu, D. et al. Understanding the relationship between China's eco-environmental quality and urbanization using multisource remote sensing data. *Remote Sens.* **14**, 198 (2022).
59. Lin, G. C. S. Reproducing spaces of Chinese urbanisation: new city-based and land-centred urban transformation. *Urban Stud. Routledge* **44**, 1827–1855 (2007).
60. Tan, K., Zhou, S., Li, E. & Du, P. Assessing the impact of urbanization on net primary productivity using multi-scale remote sensing data: a case study of Xuzhou, China. *Front. Earth Sci.* **9**, 319–329 (2015).
61. Rausch, L., Friesen, J., Altherr, L., Meck, M. & Pelz, P. A holistic concept to design optimal water supply infrastructures for informal settlements using remote sensing data. *Remote Sens.* **10**, 216 (2018).
62. Samper, J., Shelby, J. A. & Behary, D. The paradox of informal settlements revealed in an ATLAS of informality: findings from mapping growth in the most common yet unmapped forms of urbanization. *Sustainability* **12**, 9510 (2020).
63. Gevaert, C. M., Persello, C., Sliuzas, R. & Vosselman, G. Informal settlement classification using point-cloud and image-based features from UAV data. *ISPRS J. Photogramm. Remote Sens.* **125**, 225–236 (2017).
64. Addo, K. A. Urban and peri-urban agriculture in developing countries studied using remote sensing and in situ methods. *Remote Sens.* **2**, 497–513 (2010).
65. Brown, M. E. & McCarty, J. L. Is remote sensing useful for finding and monitoring urban farms? *Appl. Geogr.* **80**, 23–33 (2017).
66. Ramezan, C. A., Warner, T. A., Maxwell, A. E. & Price, B. S. Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. *Remote Sens.* **13**, 368 (2021).
67. Gislason, P. O., Benediktsson, J. A. & Sveinsson, J. R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **27**, 294–300 (2006).
68. European Space Agency. Sentinel-2 MSI level-1C TOA reflectance. *Sentinel Online* https://doi.org/10.5270/S2_-742ikth (2022).
69. Gorelick, N. et al. Google Earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).

70. Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8**, 127–150 (1979).
71. Huete, A. R., Liu, H. Q., Batchily, K. & van Leeuwen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* **59**, 440–451 (1997).
72. Diek, S., Fornallaz, F., Schaepman, M. E. & De Jong, R. Barest pixel composite for agricultural areas using Landsat time series. *Remote Sens.* **9**, 1245 (2017).
73. McFeeters, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **17**, 1425–1432 (1996).
74. Liou, Y.-A., Le, M. S. & Chien, H. Normalized Difference Latent Heat Index for remote sensing of land surface energy fluxes. *IEEE Trans. Geosci. Remote Sens.* **57**, 1423–1433 (2019).
75. Lebourgeois, V. et al. A combined Random Forest and OBIA Classification Scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sens.* **9**, 259 (2017).
76. Tassi, A. & Vizzari, M. Object-oriented LULC classification in Google Earth engine combining SNIC, GLCM, and machine learning algorithms. *Remote Sens.* **12**, 3776 (2020).
77. Achanta, R. & Süsstrunk, S. Superpixels and polygons using simple non-iterative clustering. In *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 4895–4904 (IEEE, 2017).
78. Tuzcu Kokal, A., Sunar, A. F., Dervisoglu, A. & Berberoglu, S. The use of spectral and textural features in crop type mapping using Sentinel-2a images: a case study, Çukurova Region, Turkey. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLIII-B3-2021**, 117–122 (2021).
79. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973).
80. Conners, R. W., Trivedi, M. M. & Harlow, C. A. Segmentation of a high-resolution urban scene using texture operators. *Comput. Vis. Graph. Image Process.* **25**, 273–310 (1984).
81. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **374**, 20150202 (2016).

## Acknowledgements

The research in this paper is based research conducted as part of a taught postgraduate degree at the School of Geography and the Environment at the University of Oxford, as well as a design teaching fellowship at the Department of Architecture, Art and Planning at Cornell University. The author thanks A. Lora-Wainwright, C. Rossi and A. Vasudevan for their insightful comments.

## Author contributions

H.W. confirms sole responsibility for study conception and design, data collection, analysis and interpretation of results and manuscript preparation.

## Competing interests

The author declares no competing interests.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s44284-024-00038-4.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s44284-024-00038-4.

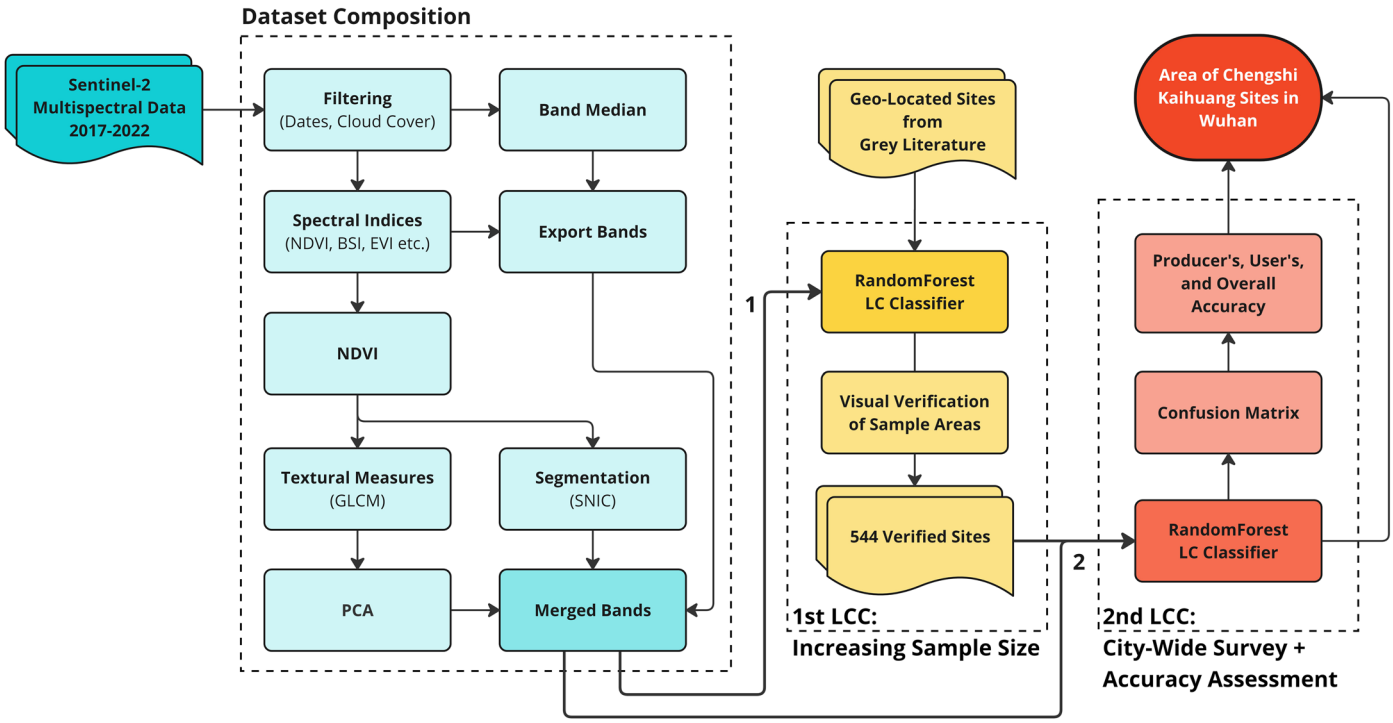
**Correspondence and requests for materials** should be addressed to Hanxi Wang.

**Peer review information** *Nature Cities* thanks Jamie Gillen, Susana Toboso-Chavero and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s44284-024-00038-4


**Extended Data Fig. 1 | Typical demolition process of rapid urbanization in Wuhan.** Counterclockwise from top left, the satellite images, denoted by year, show the typical land use changes of an area undergoing development: from rural village in 2000, through rapid demolition and construction, to new urban district in 2022. The occurrence of vacant lots in the course of urbanization become likely sites for CK, see red dashed areas in 2016 and 2022. Maps data: Google, © 2023 Maxar Technologies.

https://doi.org/10.1038/s44284-024-00038-4




**Extended Data Fig. 2 | Flowchart of the supervised machine learning land cover classification design.** Diagram illustrating step-by-step processes of the land cover classification carried out for CK detection.

https://doi.org/10.1038/s44284-024-00038-4

**Extended Data Table 1 | Dates of VHR satellite data used from Google Earth for visual assessment**

| Year | Acquisition Date (mm-dd) | Image Sources |
|---|---|---|
| 2000 | 9-21/11-1/12-23 | Landsat / Copernicus/Maxar Technologies |
| 2002 | 6-7/8-5 | Maxar Technologies |
| 2003 | 2-19/3-27 | Landsat / Copernicus/Maxar Technologies |
| 2004 | 2-27 | Landsat / Copernicus/Maxar Technologies |
| 2005 | 1-6/4-6/5-2 | Landsat / Copernicus/Maxar Technologies |
| 2006 | 1-27/12-25 | Maxar Technologies |
| 2007 | 9-21/11-27 | Maxar Technologies |
| 2008 | 5-20 | Maxar Technologies |
| 2009 | 1-22/7-16/12-22 | Maxar Technologies |
| 2010 | 1-2/10-16 | Maxar Technologies |
| 2011 | 7-29 | Maxar Technologies |
| 2012 | 4-26/7-25/9-14/11-2 | Maxar Technologies |
| 2013 | 2-10/3-29/6-13/7-24/8-16/9-29/10-10 | CNES / Airbus/Maxar Technologies |
| 2014 | 1-2/7-21/8-5/10-23/12-6 | Maxar Technologies |
| 2015 | 1-21/7-29/8-30/10-25 | CNES / Airbus/Maxar Technologies |
| 2016 | 2-20/6-16/7-22/8-14/11-6/12-27 | CNES / Airbus/Maxar Technologies |
| 2017 | 1-1/2-13/2-24/3-2/5-5/5-28/6-2/11-1/12-9 | CNES / Airbus/Maxar Technologies |
| 2018 | 4-19/6-15/9-16/10-18 | CNES / Airbus/Maxar Technologies |
| 2019 | 4-17/6-14/10-17/12-26 | CNES / Airbus/Maxar Technologies |
| 2020 | 2-25/10-30 | Maxar Technologies |
| 2021 | 2-21/3-24/7-22 | CNES / Airbus/Maxar Technologies |
| 2022 | 4-12/5-?? | Maxar Technologies |

https://doi.org/10.1038/s44284-024-00038-4

**Extended Data Table 2 | Landcover classes for supervised classification**

| LCC Classes | Data Source | Description |
|---|---|---|
| CK | Google Earth | Chengshi kaihuang sites |
| Non-CK | OpenStreetMap | Includes a range of existing land uses that are easily confused for chengshi kaihuang: trees, grassland, commercial farms, sports fields, wetland, bare soil; as well as a few typical land uses: buildings, roads, water bodies. |

https://doi.org/10.1038/s44284-024-00038-4

**Extended Data Table 3 | Spectral indices used in machine learning landcover classification**

| Type | Indices | Bands Used (Sentinel-2) | Formula | Source |
|---|---|---|---|---|
| Vegetation | NDVI | NIR/R | $(NIR-R)/(NIR+R)$ | 70 |
| Vegetation | TCARI/ OSAVI | R/G/RE1/NIR | $\frac{3(\Re 1-R)-0.2(\Re 1-G)(\Re 1/R)}{1.6(NIR-R)/(NIR+R+0.16)}$ | 82 |
| Vegetation | EVI | R/B/NIR | $\frac{2.5(NIR-R)}{NIR+6R-7.5B+1}$ | 71 |
| Bare Soil | BSI | R/B/NIR/SWIR1 | $\frac{(R+SWIR1)-(NIR+B)}{(R+SWIR1)+(NIR+B)}$ | 72 |
| Water | NDWI | G/NIR | $\frac{G-NIR}{G+NIR}$ | 73 |
| Heat | NDLI | G/R/SWIR1 | $\frac{G-R}{G+R+SWIR1}$ | 74 |

https://doi.org/10.1038/s44284-024-00038-4

**Extended Data Table 4 | Additional bands for classifier training**

| Type | Indices | Source |
|---|---|---|
| Textural Measures | GLCM | [79] |
| Object-Based Image Segmentation | OBIA | [77] |
| Demolition History | BSI_SD | |

https://doi.org/10.1038/s44284-024-00038-4

**Extended Data Table 5 | Classifier datasets, consumer accuracies, and producer accuracies for 2017–2022**

| | YEAR | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|---|
| INITIAL LCC RESULTS | | | | | | | |
| CHENGSHI KAIHUANG | Training + Validation Pts | 70 | 70 | 70 | 100 | 100 | 100 |
| | Consumer accuracy | 85.0% | 87.0% | 87.5% | 90.9% | 91.7% | 95.5% |
| | Producer Accuracy | 81.0% | 69.0% | 63.6% | 71.4% | 88.0% | 70.0% |
| NON-CHENGSHI KAIHUANG | Training + Validation Pts | 700 | 700 | 700 | 1000 | 1000 | 1000 |
| | Consumer accuracy | 98.0% | 95.5% | 98.0% | 97.2% | 98.9% | 96.8% |
| | Producer Accuracy | 98.0% | 98.5% | 99.5% | 99.3% | 99.3% | 99.6% |
| FINAL LCC RESULTS | | | | | | | |
| CHENGSHI KAIHUANG | Training + Validation Pts | 504 | 360 | 410 | 636 | 718 | 880 |
| | Consumer accuracy | 93.6% | 95.8% | 92.3% | 89.2% | 93.1% | 94.5% |
| | Producer Accuracy | 75.0% | 62.2% | 76.4% | 79.7% | 72.0% | 79.7% |
| NON-CHENGSHI KAIHUANG | Training + Validation Pts | 11,000 | 13,000 | 11,000 | 11,000 | 11,000 | 17,000 |
| | Consumer accuracy | 98.8% | 98.9% | 99.2% | 98.8% | 98.0% | 99.0% |
| | Producer Accuracy | 99.7% | 99.9% | 99.8% | 98.45% | 99.6% | 99.8% |

nature portfolio

Corresponding author(s): DBPR, NATCITIES-23090385C

Last updated by author(s): Jan 17, 2024

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | The remote sensing data leveraged in the study were accessed via the Google Earth Engine Code Editor platform. High-resolution historical satellite/aerial imagery were accessed through Google Earth Pro's desktop application. No software was used for the collection of gray literature, which were accessed and analyzed manually. |
| Data analysis | Remote sensing analyses conducted in this study were carried out using Google Earth Engine (GEE) Code Editor platform. All equations and methods leveraged in the code for the supervised machine-learning land cover classification are described in the Methods section. The code included the use of algorithms provided by the GEE platform: the RandomForest algorithm ("ee.Classifier.smileRandomForest"), the SNIC image segmentation algorithm ("ee.Algorithms.Image.Segmentation.SNIC"), as well as established tools such as the GLCM textural measure computation ("ee.Image.glcmTexture") and the Eigen analysis array for Principal Component Analysis ("https://developers.google.com/earth-engine/guides/arrays_eigen_analysis"). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

This study leverages publicly available satellite data from the European Space Agency and Google Earth Pro. Other data collected and generated in this study have been uploaded to https://figshare.com/projects/Cities_Gone_Wild_Understanding_Informal_Ruralization_Within_China_s_Rapid_Urbanization_2023-12-14_/189495. This includes the rasterized results of the supervised land cover classification for chengshi kaihuang sites; a spreadsheet of the land use histories of sampled chengshi kaihuang sites; and a spreadsheet listing national occurrences of chengshi kaihuang collected from grey literature (social media posts by non-public entities are not included to protect individual anonymity).

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | This study did not involve human research participants. |
| Population characteristics | This study did not involve human research participants. |
| Recruitment | This study did not involve human research participants. |
| Ethics oversight | This study did not involve human research participants. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | The study is a remote sensing study of an informal spatial phenomenon of urban ruralization, called chengshi kaihuang, in the central districts of Wuhan, China. Leveraging publicly available remote sensing data, the study used a shallow machine-learning algorithm to perform supervised land cover classification (LCC) and identify chengshi kaihuang sites at the urban scale. Subsequent analysis of land use histories before and after emergence were conducted to understand the characteristics of chengshi kaihuang and its relationship with the city's rapid urbanization. |
| Research sample | Samples used to conduct the study consisted of data points generated from the open access land use/land cover (LU/LC) data on OpenStreetMap and seven sites of chengshi kaihuang geolocated through gray literature. |
| Sampling strategy | The strategy for collecting chengshi kaihuang samples relied upon searching for specific keywords and hashtags on China's social media platforms. The number of chengshi kaihuang (CK) sites used to generate sample points for the remote sensing analyses in this study was determined by how many could be geo-located and verified to be true, sufficiently large (for Sentinel-2's 10m resolution) and temporally stable (must persist longer than a year). The initial LCC then yielded 544 additional CK sites that could be used to provide a much larger sample for the final, more accurate, LCC. The types of LU/LC collected for the non-CK sample included a range of existing land uses that are easily confused for chengshi kaihuang: trees, grassland, commercial farms, sports fields, wetland, bare soil; as well as a few typical land uses: buildings, roads, water bodies. |
| Data collection | LU/LC regions were downloaded from OpenStreetMap via QGIS. Verified chengshi kaihuang regions were marked out using Google Earth Pro. Both the CK and non-CK samples were uploaded as shapefile layers onto GEE, where data points were randomly generated within the regions for the LCC. |
| Timing and spatial scale | The analysis was conducted for each year between 2017 and 2022. The study area included the central urban districts of Wuhan, i.e. Jiang'an, Jiangshan, Qiaokou, Hanyang, Wuchang, Qingshan and Hongshan. |

| | |
|---|---|
| Data exclusions | No collected data was excluded from analysis. |
| Reproducibility | The code for each analyses was ran multiple times to verify their reproducibility. |
| Randomization | The training and verification points were generated randomly using Google Earth Engine's ee.FeatureCollection.randomPoints function. |
| Blinding | Blinding is not relevant as this is a remote sensing study where there are no participants that can be influenced. |

Did the study involve field work? ☐ Yes ☒ No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |