RESEARCH ARTICLE

ADVANCED SUSTAINABLE SYSTEMS
www.advsustainsys.com

Cities and Urbanization: Balancing the Environmental and Socioeconomic Dimensions of Sustainability

Becky P.Y. Loo, Ka Ho Tsoi, Xinghua Feng, Hongsheng Zhang, Yinyi Lin, Zhiran Huang, Raffaele Lafortezza, Zhenci Xu, and Hui Lin*

With an increasingly urbanized world, there is an urgent need to examine how cities may evolve and achieve sustainability. This paper systematically looks at the Greater Bay Area (GBA) and the Poyang Lake Region (PLR) in China to examine the spatial processes for insights into cities and urbanization, balancing the environmental and socio-economic dimensions. A total of 226 805 cells are analyzed to unveil the relationship between sustainability changes in 2015–2019 period and urban form indicators, considering sociodemographic variables, geographical features, and city size as control variables. Two tree-based machine learning models (Random Forest and XGBoost) are developed. This study provides evidence that a monocentric urban form and a high share of small activity clusters are not good for sustainability. For each urban form indicator, there is a non-linear relationship with sustainability. The results of the machine learning models reconfirm the sustainability benefits of having a strong second activity cluster comparable to the largest one. When planning cities, some forms of land use buffering are desirable. There is also support for developing relatively large activity nodes and promoting compactness in urban form. Beyond urban form characteristics, the levels of urbanization, economic development, and population are still highly relevant.

1. Introduction

We are living in an increasingly urbanized world. There are few regions where the natural ecosystem is largely undisturbed by human settlements. Some examples are part of the Amazon region and Antarctica.[1,2] Apart from trying to protect these regions, we urgently need to examine how cities may evolve and achieve sustainability by maximizing the synergy between urban areas and the surrounding rural areas. On the one hand, nature is to be preserved as much as possible by minimizing environmental damages and threats of increasing urbanization. On the other hand, nature serves to support the human race, whose wellbeing is an essential component of overall sustainability. With decades of research about sustainability, it is believed that there are several key elements of sustainability, including the three pillars of the environment, economy, and society,[3–6] as well as the time dimension, suggesting the importance of resilience.[4,7,8] With decoupling– a process that economic growth is no longer positively linked to negative environmental and social externalities, it is possible to achieve the simultaneous improvement of environmental, economic, and social indicators.[9–11] There is a fine balance that sustainable systems should strive to achieve.

B. P. Loo, X. Feng, H. Lin
School of Geography and Environment
Jiangxi Normal University
Nanchang 330022, China

B. P. Loo, K. H. Tsoi, H. Zhang, Y. Lin, Z. Xu
Department of Geography
The University of Hong Kong
Hong Kong, China
E-mail: jasonf1d@connect.hku.hk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adsu.202300401

DOI: 10.1002/adsu.202300401

Z. Huang
Department of Building and Real Estate
The Hong Kong Polytechnic University
Hong Kong, China

R. Lafortezza
Department of Soil, Plant and Food Sciences
University of Bari
Bari 70126, Italy

R. Lafortezza
Research Center for Urban Forestry
Beijing Forestry University
Beijing 100107, China

In history, cities have been seen as destroying nature and are detrimental to natural ecosystems directly through the replacement of natural habitats with urban land development and indirectly through proximate impact.[12] Examples include the clearing of vegetation for various forms of buildings and infrastructure, an increase in the share of impervious surfaces (e.g., roads), the destruction of animal habitats, and the pollution of air, water, and land due to industrialization and urban wastes. An often-overlooked aspect is that, in the process of urbanization, nature has been supporting a growing human population. From 2010 to 2021, the global population has increased from 6.99 to 7.91 billion.[13] The Earth's land surface is more or less fixed. To cater to the increased global population, the spatial alternative to world urbanization is to spread the increased population to live in wider rural areas with potentially more profound environmental knock-on effects. Hence, the sustainability impact of cities needs to be understood beyond the urban areas and with respect to the rural areas as well.

Spatially, we argue that comprehensive sustainability needs to be considered at the city-region scale, balancing the spatiality of various environmental, economic, and social considerations. Cities nowadays typically administer both urban and rural areas. In China, for example, the "city leading counties" system suggests that city jurisdictions already encompass nearby rural towns and counties. Within functional regions consisting of multiple cities, there can be different forms of spatial complementarity and trade-offs across urban and rural areas.[14,15] Considering sustainability at the regional scale allows for various feedback loops of the human-environment interaction system beyond city boundaries. Under the challenges of climate change, cities with a better natural ecosystem have been more successful in combating urban heat island effects.[12,16–18] Conversely, cities can foster a positive peri-coupling loop by investing in environmental conservation, education, and research, as well as legislating for the protection of physical landscapes, flora and fauna within the region.[19,20]

Moreover, the spatial pattern of human settlements and activities within cities matters. Tracing through the urban science literature, researchers have been using different methods to describe these morphological and functional characteristics as "urban form." [21] The two extremes of urban characteristics are compact cities and urban sprawl, but our understanding of the city's evolutionary process is still very limited.[22,23] Empirical evidence suggests that the agglomeration of people and activities can lead to a more efficient use of space and an increase in economic productivity that helps to conserve natural ecosystems surrounding cities.[24,25] Yet, as cities become larger, primacy and a monocentric urban form can give rise to traffic congestion and other negative environmental and social externalities.[26,27] Hence, a type of centralized decentralization or polycentricity is desirable. Polycentricity of cities generally refers to a more balanced distribution of urban centers with similar importance.[28,29] Environmentally, it is envisioned that urban compactness and polycentricity alleviate traffic-related air pollution due to shorter commuting.[30] Economically, denser urban areas can stimulate local services and business, enhance the diversity of employers and thus job opportunities, and rejuvenate the city centers.[24] Regarding social equity, socioeconomic segregations are expected to be reduced by providing higher accessibility to different types of facilities.[31]

While research suggests that compact and polycentric city development can be a more sustainable approach to urban development, there is no literature that examines whether compact cities are good for sustainability beyond the narrowly defined city boundaries and within functional regions. Such information is urgently required because of the rapid rate of worldwide urbanization.[32] In particular, urbanization in China is notable both due to its unprecedented rate and the massive spatial extent.[33] The process has been happening most rapidly in major regions of strategic importance in the country.[34,35] Hence, this study examines a core scientific question– Are compact cities good or bad for sustainability? In answering this question, this study adopts an innovative approach by introducing six urban form indicators to measure city compactness and polycentricity. It evaluates how urban form contributes to sustainable development at both the grid and city-region levels that encompass both urban and rural areas. Two major regions located in different parts of China (one along the coast and one in the interior) are chosen to examine the spatial processes of urban development for insights and lessons of developing cities that are more conducive to comprehensive sustainability, balancing the environmental and socio-economic dimensions. It provides empirical, though preliminary, evidence about whether the process of agglomerating people and activities in compact urbanized areas is detrimental or beneficial to sustainability.

2. Results

2.1. Environmental and Socioeconomic Dynamics

Figure 1 shows the spatial patterns of sustainability changes (*SDS*) in the Greater Bay Area (GBA) and the Poyang Lake Region (PLR), with the municipality boundaries overlay on top of the grids. Grids that only consist of virgin land without human settlements in both 2015 and 2019 are shown in blue. Within GBA, one sees that the green tone is strong in the central area and the blue tone dominates in the peripheral area of the region. The worst red shades are found in the North near the boundary of Zhaoqing (ZHQ) and Foshan (FOS). Within PLR, the blue tone is overwhelming, suggesting precious virgin land in this region. Yet, each of the eleven municipalities has sketches of orange tone, suggesting a deterioration in at least one of the sustainability indicators over the study period. Areas having clearly worsening sustainability performance (red with $SDS \leq -2$) are found in Pingxiang (PIX) and Nanchang (NAC).

Figure 2 shows a summary of the changes at the grid level in the GBA and PLR. For the former, ≈32.8% of the cells (18.6% + 8.4% + 5.8%) have positive scores of improvements in at least one sustainability indicator over the study period. The grids (10.4%) suffered from a decline of at least one sustainability dimension. The 10.9% grids with human settlements have no change. In PLR, ≈5.3% of the cells have negative scores of lower sustainability and 3.3% of cells have positive scores of improvements in comprehensive sustainability. About 91.7% are having no major changes, including areas with human settlements (5.7%) and those without (85.7%).

Next, we briefly describe the *SDS* pattern at the city level within each region. Detailed changes in *SDS* for each city are shown in Figure S1 of Appendix A (Supporting Information). Focusing

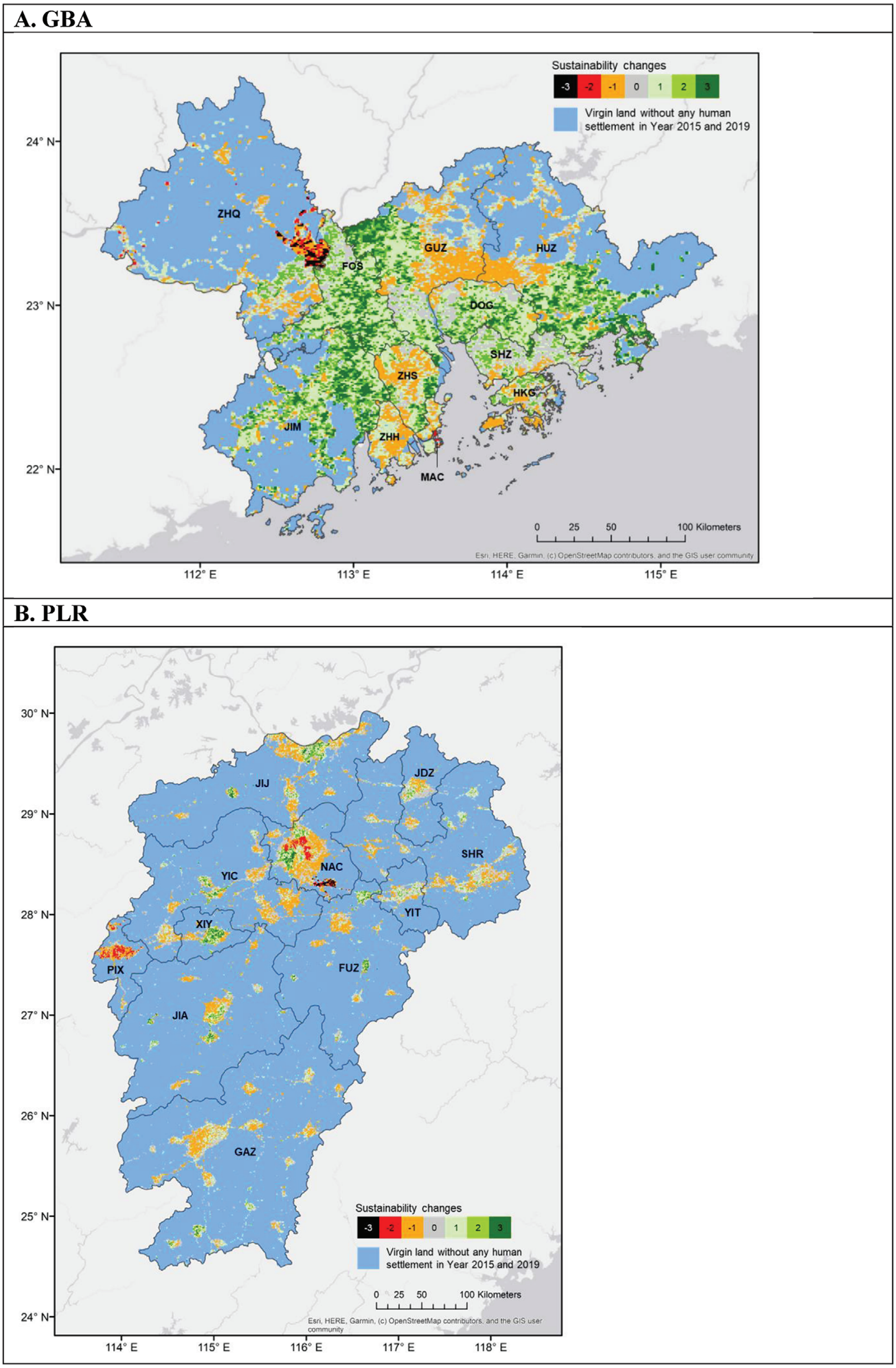

Figure 1. Spatial patterns of SDS in GBA and PLR.

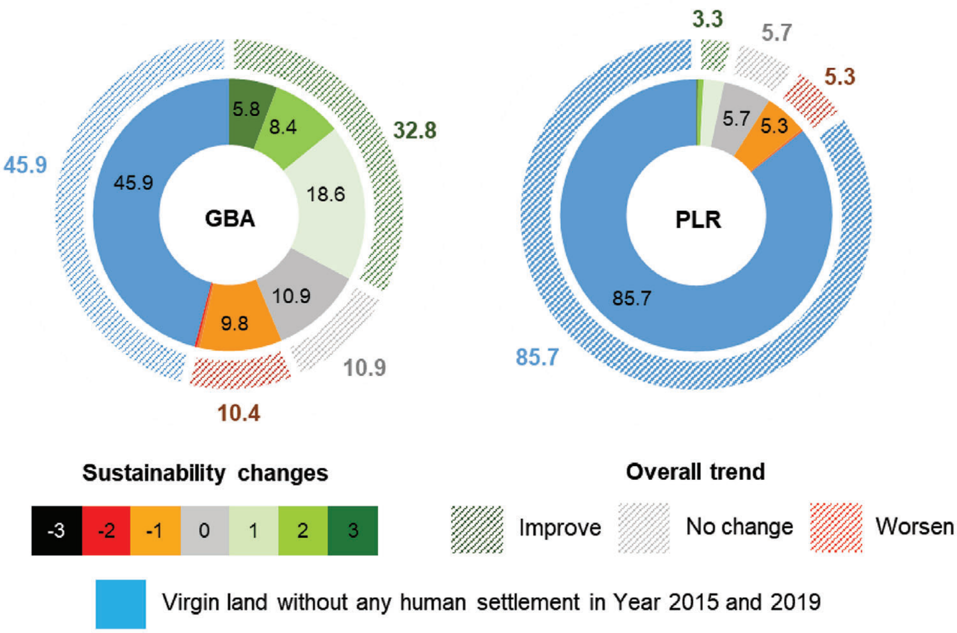

Figure 2. Summary of changes of SDS at the grid level in GBA and PLR.

on the highest value of *SDS* of 3 that all sustainability indicators have improved, the best-performing cities in GBA are Foshan (15.2%), Dongguan (10.8%), and Jiangmen (9.0%). In contrast, Macau (51.7%), Zhuhai (34.5%), and Zhongshan (32.4%) have the worst performance with the highest share of negative *SDS* scores. Within PLR, the performances are less variable. Only Xinyu has areas achieving a value of *SDS* of three in this period. The best-performing cities with the highest share of positive *SDS* are Xinyu (11.3%), Nanchang (9.4%), and Yingtan (5.8%). In contrast, those cities with negative *SDS* or lower sustainability are Nanchang (20.9%), Pingxiang (15.3%), and Yichun (7.0%). In other words, Nanchang has a notably high variability of sustainability performance over its city jurisdiction. Under the "city leading counties" system, Nanchang administers a number of districts and counties. Its administered Jinxian County, for instance, has been experiencing an absolute reduction in population and poorer economic performance during the study period, while Xinjian District closer to the urban district has developed from a county to a new activity node with better facilities and infrastructure, leading to higher *SDS* scores.

2.2. Relationship Between Sustainability Dynamics and Urban Form Characteristics

Taking cities as study units, what is the relationship of City-level SDS (*City-SDS*) with the urban form indicators? **Figure 3** shows the correlation plots of *City-SDS* with the six urban form (UF) indicators. We see that the environmental sustainability change in the Enhanced Vegetation Index (EVI) (*City-SD1*) has a statistically significant correlation with *UF2* (correlation coefficient = −0.53; $p < 0.05$) and *UF3* (correlation coefficient = 0.66; $p < 0.01$). The results suggest a positive relationship between a concentrated urban form (*UF3*) and a negative relationship between activity

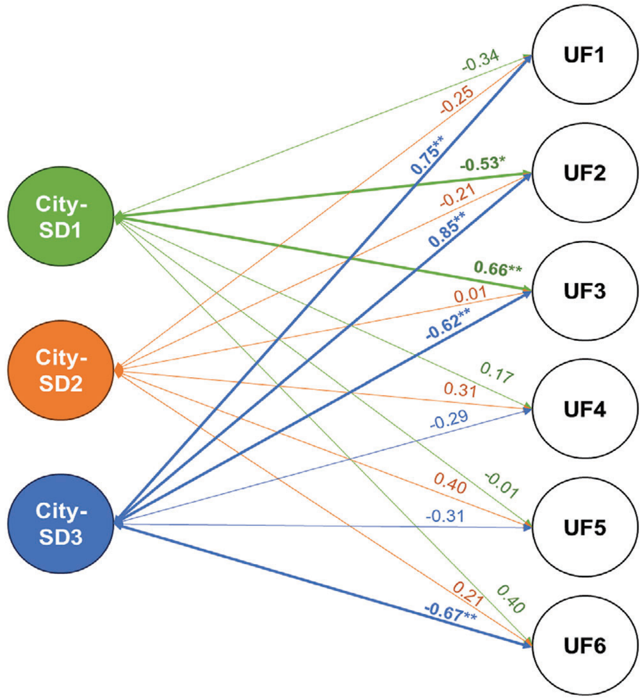

Figure 3. Summary of correlations of component city sustainability scores and six urban form indicators in 2019. Notes: [1] *City-SD1-3* are the city-level sustainability dynamic scores of *SD1–SD3*. *UF1-6* are urban form indicators 1–6.[2] The number above the line is the correlation coefficient. * is significant at 0.05 level; ** is 0.01 level.

cluster density (*UF2*) and environmental sustainability at the city level. For the social dimension in the mortality rates per 1 000 persons (*City-SD2*), the correlations with all urban form indicators are not strong and are not statistically significant at 0.05 level. In contrast, the correlations of the economic dimension in the Gross Domestic Product (GDP) per capita (*City-SD3*) and various urban form indicators are much stronger. The correlations of *City-SD3* with all urban form indicators, except *UF4* and UF5, are statistically significant at 0.05 level. The correlations with *UF2* (correlation coefficient = 0.85; $p < 0.01$) and *UF1* (correlation coefficient = 0.75; $p < 0.01$) are the strongest. The finding indicates that economic sustainability relies heavily on the density of activity clusters per million population. Also, more substantive activity clusters with a size larger than the average (*UF6*: correlation coefficient = −0.67; $p < 0.01$) and a wider spatial spread of activity clusters (*UF3*: correlation coefficient = −0.62; $p < 0.01$) correlate with improvements in economic sustainability.

Figure 4 shows the scatterplots of *City-SDS* and the six urban form indicators in 2019. In each scatterplot, there are 22 observations, corresponding to the municipalities in two regions. Where clear outliers are detected, they are removed. Linear and polynomial lines have been tried for the best-fit lines. Power 2 is chosen based on the highest R-squares achieved. We have also examined the scatterplots for *City-SDS* and changes in the urban form indicators from 2015 to 2019 (Figure S2 of Appendix A, Supporting Information). The picture is generally the same. As changes in urban form indicators need to be interpreted with absolute levels achieved, we focus on analyzing the UF values that are easier to interpret. Also, a one-way analysis of variance (ANOVA) was conducted to examine whether there are significant differences between *City-SDS* and the six urban form indicators (*UF1*-6) in GBA and PLR. The results are summarized in Table S4 of Appendix A (Supporting Information). Generally, the values of *City-SDS* in GBA (M = 0.72, SD = 0.54) are significantly higher than those in PLR (M = −0.05, SD = 0.12) ($F(1, 16) = 17.1$, $p = 0.00$). In other words, the sustainability performance of the two regions differs significantly at 0.05 level, with cities in the GBA being more sustainable. However, all urban form indicators do not show significantly differences among cities in GBA and PLR at a 0.05 significance level. They show substantial variations both within and across the two regions.

Focusing on *UF1* (density of activity clusters per million population by area), there is an invert-U relationship in GBA, implying that lower compactness or urban expansion (higher *UF1*) is positively related to sustainability at first. After the critical threshold at ≈0.5–0.6, higher compactness (lower *UF1*) is positively related to comprehensive sustainability (*SDS* on the y-axis). For cities to be vibrant and sustainable, the development of some viable activity clusters is necessary. However, when *UF1* becomes high (>0.5–0.6), a less compact urban form (higher *UF1*) is associated with worsening sustainability performance. For PLR, the correlation seems to be weak, with most observations staying close to the x-axis. The values of *City-SDS* among cities in PLR are much less variable during the study period.

For *UF2* (density of activity clusters per million population by number), the curve in the GBA is similar to *UF1*. However, as higher values of *UF1* suggest lower compactness but higher values of *UF2* suggest more dispersion (i.e., more activity clusters per million population), the interpretation is different. Here, activity node dispersion is first positively correlated with improvement in sustainability and then it turns negative. This is a good reminder to policymakers and urban planners that polycentricity needs to be balanced carefully to ensure centralized decentralization and to avoid turning polycentricity into urban sprawl. In PLR, the trend is much weaker with almost all observations staying near the x-axis. For *UF3* (Global Moran's I), there is again a non-linear relationship. As shown in Figure 4, a higher degree of spatial concentration (higher *UF3*) first improves sustainability and then reduces it. The turning point seems to be ≈0.7. In PLR, though the curve is downward-sloping, the differences are relatively minor with the points clustered along the x-axis. For *UF1* to *UF3*, the R^2 value is generally low (ranging from 0.02 to 0.54).

For *UF4* (ratio between the second largest and the largest activity clusters), the relationship is much clearer and stronger in GBA ($R^2 = 0.67$), implying a generally upward-sloping curve with a more viable second largest activity cluster being associated with better sustainability performance. Again, similar to *UF2*, the situation in PLR is not clear with almost all observations staying near the x-axis. For *UF5* (ratio of the size of thethird largest activity cluster to the largest one in the city), the curve in GBA is upward-sloping and the R^2 is also high ($R^2 = 0.61$). In other words, a strong and viable third activity cluster is also favorable to sustainability. Finally, for *UF6* (share of activity clusters with size smaller than the average), a generally downward-sloping curve is observed in both GBA and PLR. Thus, positive changes in the share of smaller-than-average activity clusters suggest poorer overall sustainability performance.

2.3. Sustainability Dynamics and Determinants

The above analysis on sustainability dynamics and urban form does not take into account various control and confounding factors. In this section, we examine the role of urban form considering other independent variables. To recall, we build two machine-learning models to explain sustainability dynamics at the grid level with various grid-level variables (including population, GDP, economic structure, and road network density) and city-level indicators (consisting of both urban form and non-urban form variables) (Appendix A, Supporting Information). Moreover, the results of the hyperparameter tuning are also shown in Table S5 of Appendix A (Supporting Information). **Table 1** compares the model fit of the two models. We conducted the analysis for each region separately and by combining data from both regions. As our aim is to generalize and to draw scientific conclusions, results of the overall model combining all data and including the dummy regional variable (i.e., GBA = 1 or 0) are discussed. From Table 1, the performances of random forest (RF) and XG Boost (XGB) are similar. XGB has a slightly better model fit with lower mean absolute error (MAE) and mean square error (MSE), as well as higher R^2. The R^2 for the combined dataset is comparable to those of the two regions separately.

As the model results of RF and XGB are similar, we show the permutation importance of the explanatory variables for both models in **Figure 5**. The descending order of importance is arranged based on the XGB model. The permutation importance of the top-ranking factors is discussed with insights from the detailed partial dependency plots shown in **Figure 6**. Based on the feature importance, various factors are at play in explaining sustainability changes. Among them, urban form indicators are highly relevant. As urban form is the focus of this study, urban

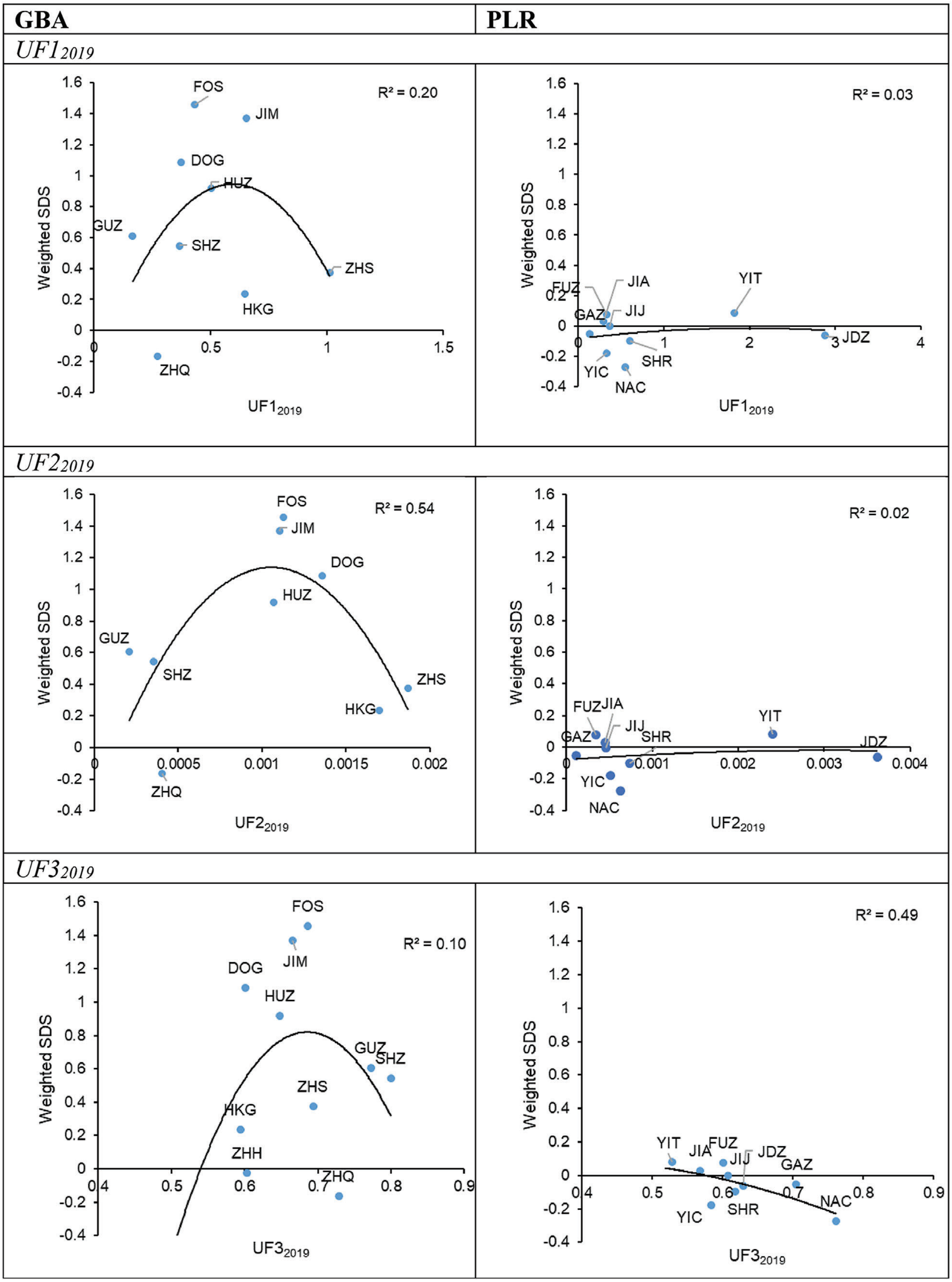

Figure 4. Scatterplots of *City-SDS* and six urban form indicators.

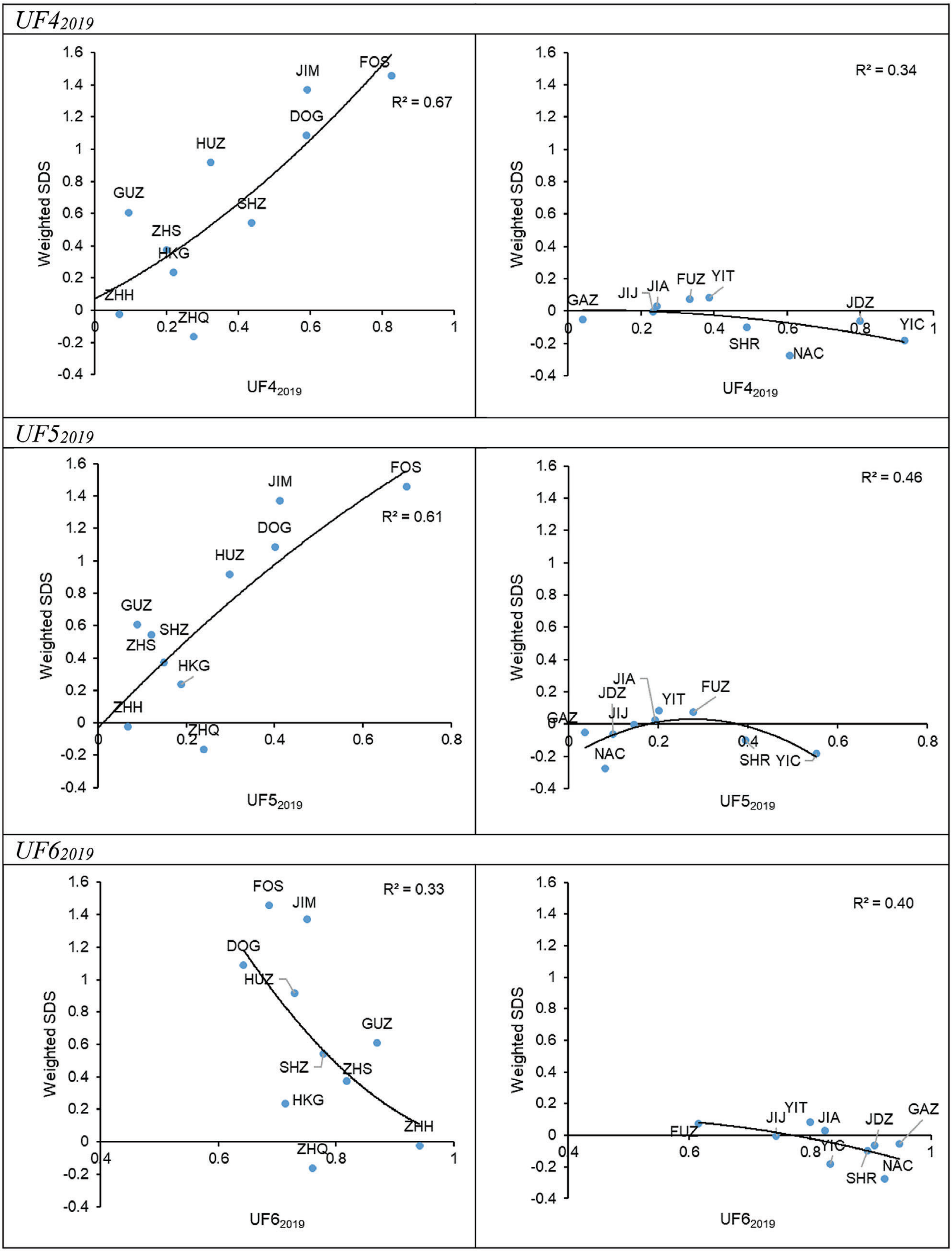

Figure 4. *Continued*

Table 1. Model fits of random forest (RF) and XG Boost (XGB) compared.

	Dataset	Random forest (RF)			XGBoost (XGB)		
		MAE	MSE	R^2	MAE	MSE	R^2
GBA	Training	0.71	0.73	0.55	0.55	0.45	0.72
	Testing	0.75	0.81	0.49	0.69	0.70	0.57
PLR	Training	0.67	0.67	0.51	0.47	0.36	0.74
	Testing	0.71	0.74	0.43	0.65	0.62	0.53
All	Training	0.73	0.77	0.53	0.53	0.43	0.73
	Testing	0.75	0.83	0.49	0.67	0.66	0.60

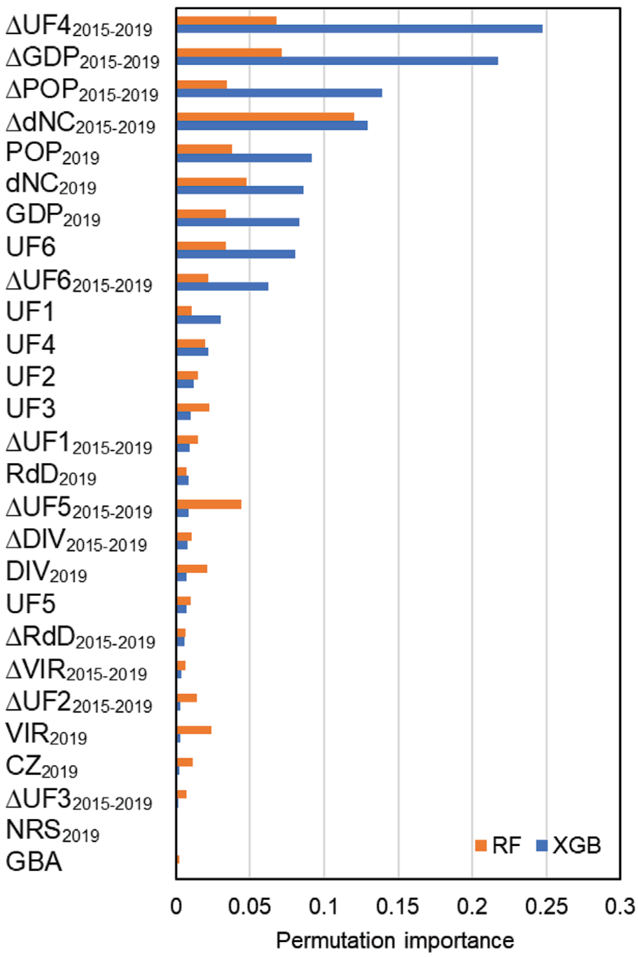

Figure 5. Permutation importance.

form variables are discussed first in descending order of importance.

$\Delta UF4_{2015\text{-}2019}$ is the most important explanatory variable. Based on the partial dependency plot, there seems to be a positive relationship that a more viable second-largest activity cluster (relative to the largest) is favorable to overall sustainability improvements. Then, the change of distance to the nearest activity cluster ($\Delta dNC_{2015\text{-}2019}$) and the distance to the nearest activity cluster in 2019 (dNC_{2019}) also matter as the fourth and sixth most important independent variables. Overall, being more distant to the nearest activity clusters suggests better overall sustainability performance. In urban planning, this finding implies that some form of buffering using, for example, natural habitat or green spaces is desirable to ensure the achievement of comprehensive sustainability. Among the top ten explanatory factors, there are three other urban form variables. They are *UF6*, $\Delta UF6_{2015\text{-}2019}$, and *UF1*. As changes in urban form indicators are difficult to interpret without respect to the absolute levels, we focus the interpretation on *UF6* and *UF1*. For *UF6*, a higher share of small activity clusters is not conducive to sustainability. For *UF1*, a larger average area of activity clusters per million population by area is not good for sustainability. The results provide evidence support for compactness in urban form as a policy goal.

Next, we turn to other variables. $\Delta GDP_{2015\text{-}2019}$, $\Delta POP_{2015\text{-}2019}$, POP_{2019}, and GDP_{2019} are among the top ten variables. This suggests that population and the level of economic development are still key explanatory factors for sustainability dynamics. From the partial dependency plots, economic growth is conducive to comprehensive sustainability. Yet, population growth is clearly exerting pressure on *SDS*. *GDP* has a weak negative relationship with *SDS* but there is hardly any change from 500 Yuan onwards at the grid level. Large population size is also having a generally negative relationship, but there is hardly any change beyond about 4000 people per grid.

3. Discussion

While the above analysis reveals how urban form can affect sustainability, it does not consider the urbanization stage of cities. Focusing on the urbanization rate (% of the total population living in urban areas) of the two regions, GBA has generally reached a mature development stage, with a high urbanization rate of 87.6% already in 2015. In 2019, it has risen slightly to 87.7%. Since the establishment of the Pearl River Delta Economic Zone in the 1990s, the urbanization rate of the region has accelerated. After decades of development, the urbanization rate has already stabilized in the recent decade. Conversely, PLR has still been under rapid changes after the establishment of the Poyang Lake Economic Region in 2010. The region's urbanization rate was 52.1% in 2015, and it has increased to 58.7% in 2019. In other words, the annual change in urbanization shows a negative relationship with the urbanization rate (Figure S3 of Appendix A, Supporting Information). The relationship is observable when GBA (R^2 = 0.31) and PLR (R^2 = 0.32) are

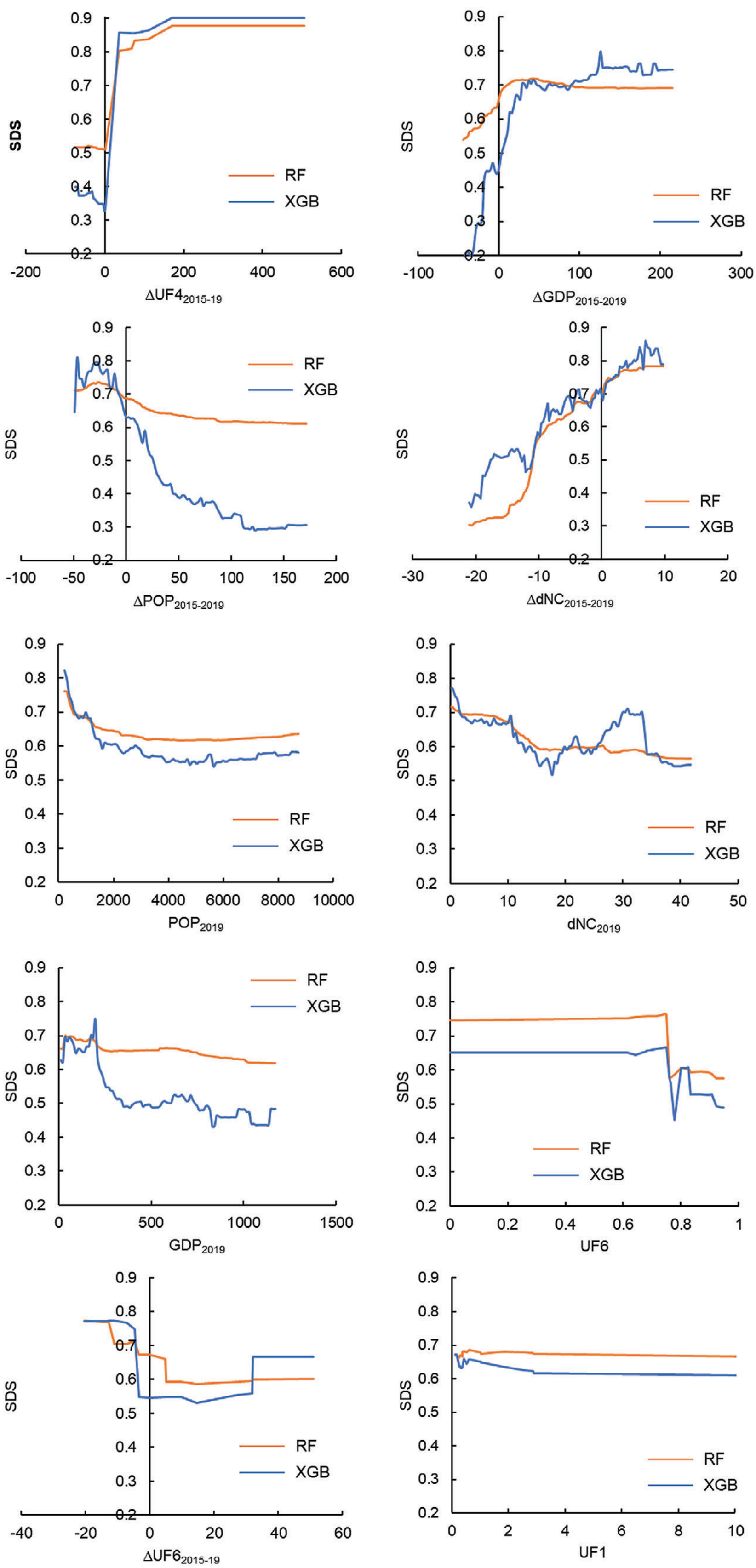

Figure 6. Partial dependency plots.

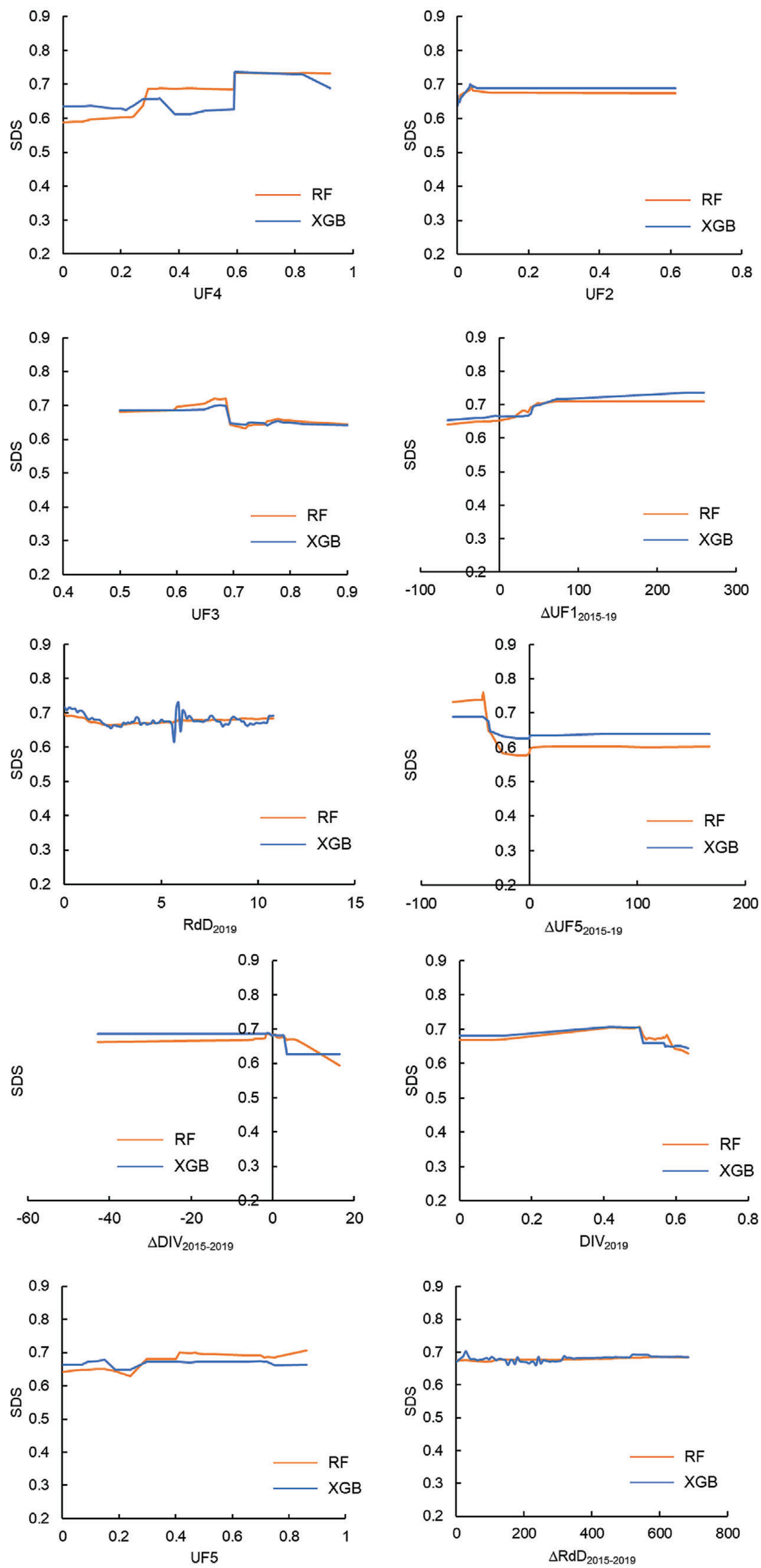

Figure 6. *Continued*

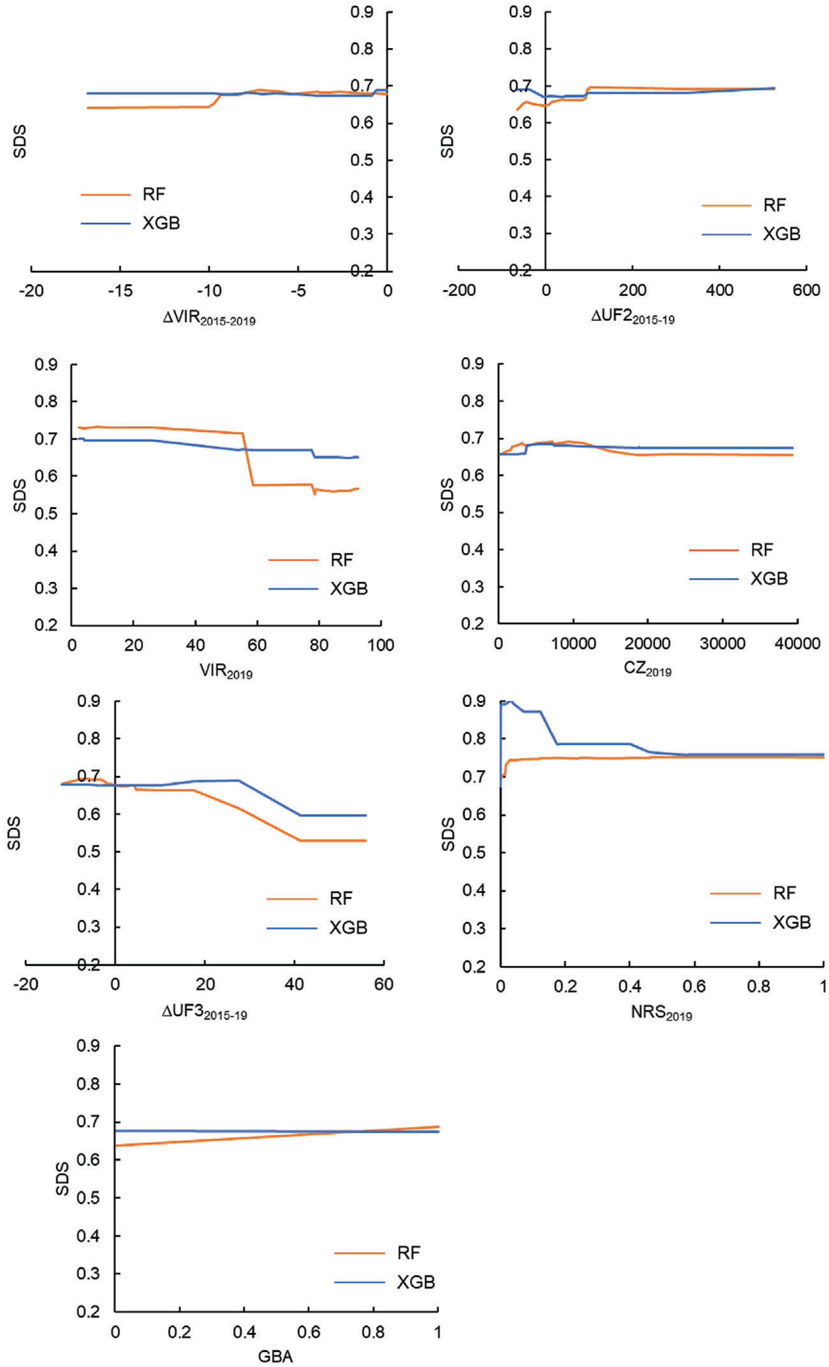

Figure 6. *Continued*

analyzed separately, and it is even stronger ($R^2 = 0.62$) when data of the two regions are combined. Further analysis of the urbanization rate and sustainability clearly shows that a more mature stage of city development (i.e., higher urbanization rate and a more gradual change of urbanization rate) is associated with better sustainability performance (higher *City-SDS*) (Figure S4 of Appendix A, Supporting Information).

Focusing on different cities within the two regions, in GBA, many cities (e.g., Hong Kong, Shenzhen, Macau, Foshan, and Dongguan) have been major population and economic centers since the 1990s (including hubs for secondary and tertiary industries). By 2015, these five (out of eleven) cities have achieved a very high level of urbanization (>90%) with well-established urban cores. In other words, about half of the cities (45.5%) in the region are highly urbanized. With a stabilized urbanization growth, most cities in GBA managed to achieve a higher degree of compactness (*UF3*) and less urban sprawl (*UF6*) (Figure S5 of Appendix A, Supporting Information). Following the establishment of GBA in 2017, Huizhou, Jiangmen, and Zhaoqing have accelerated their urbanization in the recent decade, where most of the virgin land and natural reserves are located.

In contrast, cities in PLR are generally less urbanized. Nanchang, the capital city of Jiangxi, is the most urbanized city in PLR. Many cities have undergone rapid rural transformation in the last decade after the establishment of the Poyang Lake Ecological Economic Zone in 2009 and the New Urbanization Development Plan (2014–2020). The rapid growth of urbanization among PLR cities was, unfortunately, associated with a larger extent of urban sprawl (*UF6*), reduced compactness (*UF3*), and a single dominant urban cluster with less decentralization of economic activities (*UF4*) (Figure S5 of Appendix A, Supporting Information).

The above observations are consistent with the broader literature on urbanization and sustainability, suggesting that a city's sustainability performance is partially related to its urbanization level and development stage.[14,34,35] Generally, cities undergoing rapid urbanization (i.e., a substantial increase in annual urbanization rate) face challenges of intensive land consumption for economic development.[36] This leads to significant land use changes, which often result in lower ecological and social sustainability. In contrast, cities reaching a higher maturity stage of urbanization (i.e., a lower increase in urbanization rate) tend to achieve better environmental, economic, and social sustainability performance by employing different compact city development strategies.[24,36] In particular, they include an agglomeration of higher value-added industries and more proactive environmental conservation policies to control urban expansion. Our analysis provides additional insights into the relationship between urbanization levels and sustainability by examining various urban form indicators (*UF1*-*UF6*).

As a sustainable development strategy must strike a balance between economic growth and other social and environmental concerns, the above results provide policy implications in formulating specific strategies based on the cities' development stages and spatial patterns. In GBA (or regions with high urbanization and better economic development), cities should focus more on environmental and social sustainability. In fact, GBA has started to address ecological and social concerns in the recent decade (e.g., Pearl River Delta National Forest Urban Agglomeration Construction Plan). This implies that in a more advanced state of urbanization, protecting natural areas and developing green corridors as natural buffers to prevent urban sprawl are priorities for sustainability. Maintaining the city's compactness and developing competitive activity clusters are also of prime importance. Finally, coordinated development among cities in urbanization that fosters cooperation and synergies is critical to sustainability.[37]

In regions that are still undergoing rapid urbanization and economic growth, more attention should be put on avoiding unconstrained urban sprawl by stipulating stronger land use regulations while fostering strong and vibrant activity nodes to boost economic development. The size of these activity nodes should be comparable to the main urban core for the sustainability benefits to be realized. In recent years, PLR has also adopted a series of ecological conservation policies and strengthened environmental protection regulations in wetlands, lakes and ecological areas in the past decade.[38] These can help avoid unplanned changes in land use under rapid urbanization. Nonetheless, more efforts to foster a polycentric urban form, with at least two or three largest activity nodes of comparable size within each city, are urgently needed to stop the trend towards urban sprawl.

4. Conclusion

Cities worldwide should aim at sustainability, balancing the environmental and socioeconomic dimensions. Nonetheless, characteristics of urban form have seldom been seen as variables in the sustainability "equation." [22] The problem partly lies in the lack of good indicators to generalize across different geographical contexts. With increasing world urbanization, this discussion on urban forms needs to be urgently and carefully considered to enhance sustainability. This study provides more evidence on urban form and sustainability, which is driven by the coupling between human and environmental systems within city boundaries. Cities are not homogenous built-up areas, and the ways that the development and population or activity clusters are arranged are important.

Through this study, we find evidence of improved environmental sustainability associated with higher compactness (*UF3*) and fewer activity clusters (*UF2*). The relationship of social sustainability with urban form is weak. In contrast, better economic sustainability is associated with concentrated decentralization or polycentricity, which reflects a fine balance of the density of activity clusters per million population (*UF1* and *UF2*), the proportion of activity clusters larger than the average size (*UF6*) and the spatial distribution of activity clusters (*UF3*). For each of the urban form indicators, there is a non-linear relationship with sustainability. In particular, lower compactness or urban expansion (higher *UF1*) is positively related to sustainability at first. However, the relationship reverses after the critical threshold at ≈0.5–0.6, suggesting that higher compactness (lower *UF1*) is positively related to comprehensive sustainability. For *UF2*, an increase in the number of activity clusters is first positively correlated with improvement in sustainability and then it turns negative. Then, *UF3* suggests that a higher degree of spatial concentration of activity nodes (higher *UF3*) first improves sustainability and then reduces it. The turning point seems to be ≈0.7. Besides, larger and more viable second (high *UF4*) and third (high *UF5*) largest activity clusters are beneficial to sustainability performance. Finally, a higher share of larger-than-average activity clusters (lower *UF6*) suggests improved overall sustainability performance.

Overall, polycentricity needs to be balanced carefully to ensure centralized decentralization. A monocentric urban form and too many small activity clusters are not good for comprehensive sustainability. The results of the machine learning models reconfirm the critical importance of having a stronger second activity cluster comparable to the largest one ($\Delta UF4_{2015\text{-}2019}$). When planning cities, some forms of land use buffering, such as natural habitat or green space, are desirable ($\Delta dNC_{2015\text{-}2019}$). There is also evidence for developing relatively large activity nodes (*UF4*, *UF6*, and $\Delta UF6_{2015\text{-}2019}$) and promoting compactness in urban form (*UF1*) as policy goals. Beyond urban form characteristics, the findings suggest that the level of economic development ($\Delta GDP_{2015\text{-}2019}$ *and* GDP_{2019}) and population ($\Delta POP_{2015\text{-}2019}$ *and* POP_{2019}) are still highly relevant. While economic growth is conducive to comprehensive sustainability, the pressure of population growth on sustainability is strong. At the regional level, higher sustainability has been associated with about half of the cities being in the mature stage of city development (i.e., higher urbanization rate and a more gradual change of urbanization rate). In the GBA, five out of the 11 municipalities have been highly urbanized by 2019.

However, this study has also some limitations as we only examine two major regions in China. Moreover, each dimension of sustainability is only represented by one indicator associated with one Sustainable Development Goal (SDG). After demonstrating the feasibility and value of this conceptual framework and methodology, the analysis can be widened with more sustainability indicators to encompass more SDGs at the target level. Also, future research can conduct cross-country comparisons so that the results are more generalizable. Nonetheless, the data collected and analyzed in these two regions (over 200 000 grids) are already massive, and the findings of this study provide valuable empirical insights for achieving sustainability related to SDG 15, SDG 3, and SDG 8. Last but not least, future research should be directed at examining and strengthening mutually supportive city-region interactions so that there can be a balance of all three sustainability pillars at the city-region scale.

5. Experimental Section

Study Area: This paper systematically looks at the Greater Bay Area (GBA) in South China and the Poyang Lake Region (PLR) in Central China. People living in each of the two regions shared similar local dialects and cultures, and have close economic ties as a functional regions in history.[34,39] The former, being a strategic region in national development, covers nine municipalities in Guangdong and two Special Administrative Regions (SAR). They are Dongguan, Foshan, Guangzhou, Huizhou, Jiangmen, Shenzhen, Zhaoqing, Zhongshan, Zhuhai, Hong Kong SAR, and Macau SAR. The total area is over 56 000 km^2. The latter, centers around Poyang Lake, covers eleven municipalities, including Fuzhou, Ganzhou, Jian, Jingdezhen, Jiujiang, Nanchang, Pingxiang, Shangrao, Xinyu, Yichun, and Yingtan in Jiangxi province. The total area is 166 000 km^2. Table S1 of Appendix A (Supporting Information) shows a summary of the study area and the abbreviations of city names. To properly consider the spatial dimension of cities and regions, the study area was divided into 1 km x 1 km cells. The numbers of cells were 57 913 and 168 892 for GBA and PLR, respectively. The projection systems applied in this study are WGS84_UTM49N and WGS84_UTM50N for GBA and PLR, respectively.

Sustainability Measures and Data Sources: First, indicators were selected with respect to the SDGs and the three pillars of sustainability. As a pilot study, the indicators included were limited and primarily based on data availability. This was a limitation of the study that needs to be addressed in future studies. Nonetheless, by focusing on one indicator to reflect each dimension of the environment, society, and economy, this study manages to examine the relationship between urban development (notably urban forms) and comprehensive sustainability in a more in-depth manner at fine spatial resolution. The framework can be extended to more indicators of SDGs once the feasibility and value of the approach have been established. Through considering different directions of change of SDGs, it also aims to identify whether there was any complementarity or trade-off among different dimensions of sustainability in a city region.

In relation to SDG 15 of Life on Land, the key **environmental** indicator chosen was the Enhanced Vegetation Index (EVI) (*SD1*), which indicated the level of vegetation greenness with correction factors for atmospheric conditions. Again, more specific indicators of SDG at the target level (rather than the higher goal level) were desirable at a later stage beyond this pioneer study. The data were compiled from the database of MOD13A2.006 Terra Vegetation Indices 16-Day Global 1 km (LP DAAC, 2021). Summer mean data (June, July, and August) were extracted between 2015 and 2019 on the 1 km x 1 km grid. It was measured in a range of −1–1. A larger value represents a higher level of greenness in the area.

Based on SDG 3 on Good Health and Well-being, the basic **social** indicator was the mortality rates (number per 1000 persons) (*SD2*). The mortality rate indicates the ratio of an actual number of deaths to the total population. The total population and number of deaths between 2015 and 2019 were extracted in the most disaggregated administrative unit available from government statistical yearbooks. To derive the mortality rates at a grid level, the values of both variables were distributed to each cell based on the nighttime light data using the *pro-rata* method. Specifically, the annual nighttime light data (in the unit of nW cm^{-2} sr^{-1}) between 2015 and 2019 were derived from enhanced vegetation index adjusted NTL index (EANTLI) of 2000–2013, composited NPP-VIIRS NTL data from 2012 to 2018 and DMSP-OLS radiance-calibrated NTL (RNTL) data using autoencoder network and cross-sensor calibration on Google Earth Engine and CUDA parallel computing platform.[40] Hence, the total deaths (D_i) and population (Pop_i) in grid *i* were calculated by the following equations:

$$D_i = D_y \times \frac{NL_i}{NL_y} \tag{1}$$

$$Pop_i = Pop_y \times \frac{NL_i}{NL_y} \tag{2}$$

where D_y and Pop_y indicate the number of deaths and population in administrative unit *y* respectively, NL_i refers to the nighttime light index in grid *i*, and NL_y refers to the total nighttime light index in administrative unit *y*.

In relation to SDG 8 on Decent Work and Economic Growth, the **economic** indicator was Gross Domestic Product (GDP) per capita (*SD3*), which was the ratio of actual GDP (in real 2015 Yuan) to population. Both population and real GDP were summarised in the most disaggregated administrative unit from government statistical yearbooks. Then, the value of real GDP was distributed to each grid based on nighttime light data using the *pro-rata* method. This was to reflect that economic activities were strongly correlated with the intensity of nighttime light.[41,42] In general, the real GDP in grid *i* (GDP_i) was calculated as follows:

$$GDP_i = GDP_y \times \frac{NL_i}{NL_y} \tag{3}$$

where GDP_y and NL_y indicate the GDP and nighttime light index in administrative unit *y*, and NL_i refers to the nighttime light index in grid *i*.

Urban Form Measures: The degree of urban compactness and polycentricity was the focus of this study. The concept of "natural cities" was first adopted,[43] which was a spatial model that demarcates urban clusters based on street networks and street nodes. The rationale behind this was that human activities were constrained to street nodes, and this bottom-up approach directly reflects urban aggregations with reference to human activities, rather than administrative boundaries.[43,44] Another advantage of this method was that it only used road network data which are open and could be accessed easily and measured objectively.

The spatial boundaries and road network were extracted from the Database of Global Administrative Areas and National Catalogue Service for Geographic Information.[45,46] The road network data of 2015 and 2019 were used in generating the natural cities in each city. Essentially, the aim was to capture areas of high economic density, morphological density, and mixed land use.[21,27,36,47–50] As such, natural cities were referred as activity clusters or nodes, which are more commonly known in the urban literature. This approach had been empirically tested in recent studies where the activity nodes reflected the aggregation of traffic and human activities in urban areas.[23]

Based on the activity clusters derived from the natural city model, a total of six indicators were derived. For easy reference, the descriptions of these urban form indicators are also summarized in Table S2 of Appendix A (Supporting Information). It is noteworthy that all urban form indicators are at the city (not grid) level. First, two direct indicators capture the main features of urban agglomeration. They both represented the density of activity clusters per million population. The first indicator was the **density of activity clusters per million population in terms of area (*UF1*)**. For *UF1*, a smaller value indicates a higher degree of compactness in a city since each person consumes less land area for urban development. As a monocentric compact city can give rise to heavy traffic congestion and crowding issues in a small area, it is generally considered that a form of decentralized concentration or polycentricity is desirable. The second indicator measured the **density of activity clusters per million population in terms of the number of activity nodes (*UF2*)**. A larger value over time suggested more activity nodes in the city.

The third indicator captured the degree of spatial concentration of activity clusters by the **Global Moran's I (*UF3*)** index, which measures the spatial autocorrelation according to feature values and feature locations simultaneously.[51] The equation is listed as follows:

$$I = \frac{n}{W} \times \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{4}$$

where n is the total number of features, W is the aggregate weight of all spatial weights, $w_{i,j}$ refers to the spatial weight between feature i and j, and z_i is the deviation of an attribute of feature i and its mean. The index identifies the degree of clustering of the spatial data, ranging from −1 to 1. A value closer to 1 indicates a higher degree of clustering, whereas closer to −1 represents a higher degree of dispersion, and 0 indicates a completely random spatial pattern. It is expected that a totally dispersed pattern approximating urban sprawl ($I < 0$) is not good. Generally, a positive value but below 0.75 would more closely approximate decentralized centralization or polycentricity in urban form.

The last three indicators captured the relationship of the size of activity clusters in relation to each other. According to Zipf's Law, activity clusters tend to have a heavy tail distribution, with far smaller than larger NCs.[52] The fourth indicator was the **ratio between the second-largest and the largest activity clusters (*UF4*)**. If the ratio is high, it indicates that the difference in the areas of the two largest natural cities is small, suggesting the development of dual city centers or CBDs, and hence a higher degree of polycentricity. On the contrary, a small ratio represents the dominance of one single urban center which suggests a monocentric urban form. The fifth indicator was the **ratio of the size of the third-largest activity cluster to the largest one in the city (*UF5*)**. Polycentricity is difficult to achieve in urban planning. While some cities now clearly with two urban cores (e.g., Tsim Sha Tsui and Central in Hong Kong), grooming a third viable and comparable urban core is challenging. This indicator quantifies whether the third activity cluster is similar in size to the largest. A high value suggests concentrated polycentricity with three viable urban clusters. The sixth and final indicator was **the share of activity clusters with size smaller than the average (*UF6*)**. Given the rank-size rule, it was expected that the share of smaller-than-average activity nodes would be bigger than 50%. Nonetheless, beyond this, a larger share suggests a tendency towards urban sprawl. Together, these six indicators quantify the degree of concentrated polycentricity.

At the grid level, a grid's relative position in the urban form was measured by its distance to the nearest activity cluster (*dNC*). This method was developed in a recent study.[52] In this study, the aim is to see whether the proximity of a grid to existing activity clusters is beneficial to its overall sustainability or not.

Sustainability Dynamics in the Spatial/Regional Context: Over time, changes in sustainability (*SD1* to *SD3*) were captured by integrating the dynamics of the three selected sustainability indicators. From 2015 to 2019, any change within ±5% is considered as no change and has a score of 0. The reason for including the margin of 5% for identifying changes was to recognize that statistical data in cities were often collected from various surveys that incorporate errors. Based on the general rule of statistical significance at 0.05 level, 5% was chosen to detect changes in this analysis. Thus, deterioration refers to a decline of 5% or more and was given a score of −1. An improvement of 5% or more was given a score of 1. Therefore, for each grid, the total score ranges from 3 to −3. Three is the best scenario that illustrates improvements in all three sustainability aspects, that is, an increase in EVI, a decline in mortality rates, and an increase in GDP per capita in the grid area over the study period. In contrast, −3, with deterioration in all three indicators, is the worst. This variable at the grid level is called the **Sustainability Dynamic Score (*SDS*)**. When the analysis was aggregated at the city level, the sum of *SDS* was divided by the total number of grids of the city to reflect the city size. In this way, the general sustainability dynamics could be analyzed for each city. To avoid confusion, the aggregated and weighted *SDS* at the city level is called City-level SDS (***City-SDS***).

A core objective of this study was to examine the role of urban form in sustainability dynamics at the regional level. Are city compactness and polycentricity beneficial to sustainability? To answer this, this study applies two tree-based machine learning models to analyze the relationship at the grid level. Both models consider the *SDS* in each grid as the dependent variable, and various urban form indicators ($UF1_{2019}$-$UF6_{2019}$, dNC_{2019}, and $\Delta UF1_{2015\text{-}19}$-$\Delta UF6_{2015\text{-}19}$) as independent variables. This included the key sociodemographic variables (i.e., population, GDP, economic structure, and their changes from 2015–2019) and geographical features (i.e., the ratio of activity cluster area in a grid, distance to the nearest activity cluster, road density, designated natural reserves, virgin land, and their changes), and the overall city size as control variables. The list of control variables, as well as their variable names, is shown in Table S3 of Appendix A (Supporting Information).

Machine learning models, such as Random Forest (RF) and XGBoost (XGB), are known to provide better model fit when compared with other traditional statistical models, especially when examining non-linear relationships. In running these models, hyperparameter tuning was conducted by splitting the data into both training (70%) and testing datasets (30%). A randomized fivefold cross-validation was used to derive the most suitable combinations of hyperparameters that maximize model performance. The mean absolute error (MAE), mean square error (MSE), and R^2 were used as the metrics for evaluating model performance.[53,54] Finally, permutation importance was examined to identify critical variables that were having stronger influences on the sustainability dynamics.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors would like to express their gratitude to the Ministry of Education, China and Jiangxi Normal University for the Changjiang Chair Professorship of the first author. This project is not possible without the generous support of the Changjiang Scholar scheme. In addition, the authors would also like to acknowledge the support of the RGC Postdoctoral Fellowship Scheme for the second author.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

All data used in this manuscript are publicly available from government websites and other sources listed in the references.

Keywords

cities, comprehensive sustainability, urban form

Received: August 31, 2023
Revised: February 5, 2024
Published online: March 2, 2024

[1] P. L. Ibisch, M. T. Hoffmann, S. Kreft, G. Pe'er, V. Kati, L. Biber-Freudenberger, D. A. Dellasala, M. M. Vale, P. R. Hobson, N. Selva, *Science.* **2016**, *354*, 1423.
[2] Y. Qin, X. Xiao, F. Liu, F. de Sa e Silva, Y. Shimabukuro, E. Arai, P. M. Fearnside, *Nat. Sustain.* **2023**, *6*, 295.
[3] Q. Jiang, Z. Li, S. Qu, Y. Cui, H. Zhang, Z. Xu, *Resour. Conserv. Recycl.* **2022**, *178*, 106092.
[4] B. P. Y. Loo, *Unsustainable Transport and Transition in China*, Routledge, London and New York **2018**.
[5] B. Purvis, Y. Mao, D. Robinson, *Sustain. Sci.* **2019**, *14*, 681.
[6] I. Rodríguez-Serrano, N. Caldés, C. De La Rúa, Y. Lechón, A. Garrido, *Environ. Dev. Sustain.* **2017**, *19*, 1981.
[7] S. Espiner, C. Orchiston, J. Higham, *J. Sustain. Tour.* **2017**, *25*, 1385.
[8] R. Zhao, X. Li, Y. Wang, Z. Xu, M. Xiong, Q. Jia, F. Li, *Sci. Total Environ.* **2023**, *898*, 165568.
[9] B. P. Y. Loo, D. Banister, *J. Transp. Geogr.* **2016**, *57*, 134.
[10] P. Tapio, *Transp. Policy.* **2005**, *12*, 137.
[11] K. H. Tsoi, B. P. Y. Loo, D. Banister, *Glob. Environ. Change.* **2021**, *68*, 102250.
[12] Y. Ren, R. Lafortezza, V. Giannico, G. Sanesi, Z. Zhang, C. Xu, *Sci. Total Environ.* **2023**, *80*, 163276.
[13] Our World in Data, https://ourworldindata.org/, accessed: March **2023**.
[14] Q. Li, J. Wu, C. Zhang, X. Wu, X. Wen, G. Huang, R. Lafortezza, *J. Environ. Manage.* **2022**, *303*, 114167.
[15] K. C. Seto, A. Reenberg, C. G. Boone, M. Fragkias, D. Haase, T. Langanke, P. Marcotullio, D. K. Munroe, B. Olah, D. Simon, *Proc. Natl. Acad. Sci. USA.* **2012**, *109*, 7687.
[16] B. Chen, S. Wu, Y. Song, C. Webster, B. Xu, P. Gong, *Nat. Commun.* **2022**, *13*, 4636.
[17] Y. Chen, W. Y. Chen, V. Giannico, R. Lafortezza, *Landsc. Ecol.* **2022**, *37*, 2179.
[18] E. Massaro, R. Schifanella, M. Piccardo, L. Caporaso, H. Taubenböck, A. Cescatti, G. Duveiller, *Nat. Commun.* **2023**, *14*, 2903.
[19] J. Liu, V. Hull, J. Luo, W. Yang, W. Liu, A. Viña, C. Vogt, Z. Xu, H. Yang, J. Zhang, L. An, X. Chen, S. Li, Z. Ouyang, W. Xu, H. Zhang, *Ecol. Soc.* **2015**, *20*, 44.
[20] J. Liu, *Ecol. Soc.* **2017**, *22*, 29.
[21] E. Burton, *Urban Stud.* **2000**, *37*, 1969.
[22] M. Batty, *Science.* **2008**, *319*, 769.
[23] B. P. Y. Loo, Z. Huang, *J. Transp. Geogr.* **2021**, *94*, 103108.
[24] S. E. Bibri, J. Krogstie, M. Kärrholm, *Dev. Built Environ.* **2020**, *4*, 100021.
[25] E. Conticelli, in *Sustainable Cities and Communities*, Vol. *1* (Eds: W. L. Filho, A. M. Azul, L. Brandi, P. G. Özuyar, T. Wall), Springer International Publishing, Cham **2020**.
[26] P. Gordon, H. W. Richardson, *J. Am. Plann. Assoc.* **1997**, *63*, 95.
[27] M. Neuman, *J. Plan. Educ. Res.* **2005**, *25*, 11.
[28] M. Burger, E. Meijers, *Urban Stud.* **2011**, *49*, 1127.
[29] A. Vasanen, *Urban Stud.* **2012**, *49*, 3627.
[30] J. Van Der Waals, *Tijdschr. Econ. Soc. Geogr.* **2000**, *91*, 111.
[31] E. Burton, in *Achieving Sustainable Urban Form*, Vol. *1* (Eds: K. Williams, E. Burton, M. Jenks) Spon Press, Oxon **2000**, 402.
[32] World Urbanization Prospects: The 2014 Revision, Department of Economic and Social Affairs, United Nations, New York **2015**.
[33] C. L. Gu, W. H. Guan, H. L. Liu, *Sci. China Earth Sci.* **2017**, *60*, 1067.
[34] C. Guo, F. Yang, Y. Ding, H. Liu, J. Feng, *Appl. Geogr.* **2023**, *156*, 103004.
[35] M. Yan, J. Zhao, S. Yan, M. Zhu, *Environ. Sci. Pollut. Res.* **2023**, *30*, 117082.
[36] Y. Yao, H. Pan, X. Cui, Z. Wang, *Land Use Policy.* **2022**, *115*, 106005.
[37] X. Liu, P. Guo, X. Yue, S. Zhong, X. Cao, *Ecol. Indic.* **2021**, *130*, 108056.
[38] S. Li, Z. Ying, H. Zhang, G. Ge, Q. Liu, *Chin. Geogr. Sci.* **2019**, *29*, 488.
[39] L. Hou, L. Wu, S. Ju, Z. Zhang, Y. Zhu, Z. Lai, *Growth Change.* **2021**, *52*, 1914.
[40] Z. Chen, B. Yu, C. Yang, Y. Zhou, S. Yao, X. Qian, C. Wang, B. Wu, J. Wu, *Earth Syst. Sci. Data.* **2021**, *13*, 889.
[41] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, S. Ermon, *Science.* **2016**, *353*, 790.
[42] Y. Lin, H. Zhang, H. Lin, P. E. Gamba, X. Liu, *Remote Sens. Environ.* **2020**, *242*, 111757.
[43] B. Jiang, T. Jia, *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1269.
[44] B. Jiang, Y. Miao, *Prof. Geogr.* **2015**, *67*, 295.
[45] *GDAM Maps and Data*, https://gadm.org/ accessed: August **2022**.
[46] National Catalogue Service for Geographic Information, https://www.webmap.cn/ (accessed: February **2021**)
[47] G. M. Ahfeldt, E. Pietrostefani, *The Economic Effects of Density: A Synthesis The Economic Effects of Density: A Synthesis*, Spatial Economics Research Centre, London School of Economics and Political Science, London **2017**.
[48] J. Lee, K. Kurisu, K. An, K. Hanaki, *Urban Stud.* **2015**, *52*, 1054.
[49] J. J. Lin, A. T. Yang, *Environ. Plann. B Plann. Des.* **2006**, *33*, 365.
[50] Y. Jia, L. Tang, M. Xu, X. Yang, *Ecol. Indic.* **2019**, *99*, 27.
[51] A. D. Cliff, J. K. Ord, *Spatial Processes: Models & Applications*, Pion, London **1981**.
[52] B. P. Y. Loo, Z. Huang, *Cities.* **2023**, *139*, 104380.
[53] K. Chu, A. Y. S. Lam, K. H. Tsoi, Z. Huang, B. P. Y. Loo, *IEEE Access.* **2023**, *11*, 76095.
[54] B. P. Y. Loo, Z. Fan, T. Lian, F. Zhang, *Accid. Anal. Prev.* **2023**, *185*, 107017.