Research Article

Using a Random Forest Model to Study the Impact of Local Government-Led Urbanization on Urban Sustainable Development

Yixiao Peng [iD]

School of Economics, Zhejiang University of Technology, Hangzhou 310023, Zhejiang, China

Correspondence should be addressed to Yixiao Peng; 10743@163.com

Urbanization has accelerated China's economic growth, but it has also brought many sustainability issues. This paper selected a random forest model to study the impact of local government-led urbanization on urban sustainable development. Urbanization affected urban sustainable development through government revenue expansion, land resources mismatch, and industrial structure adjustment. The results showed that the adjustment of industrial structure has the greatest impact on urban sustainable development, and the importance of the average output of industrial enterprises confirms it. Government revenue expansion and land resources mismatch are more important to the sustainable development of representative urban agglomerations. The goodness of fit of the random forest model is better than the multiple linear regression (MLR) model and the extreme gradient boosting (XGBoost) model. The generalization ability of the model is improved with the optimization of variables. The main contribution of this paper is that we have established a complete information dynamic game model on government revenue expansion, land resource mismatch, industrial structure adjustment, and urban sustainable development. And the random forest model is used to study the relationship between the above variables.

1. Introduction

In the current stage of China's urbanization process, land resource mismatch, imbalance of industrial structure, and government revenue expansion have brought negative impacts on the urban sustainable development. China's central government has regarded GDP development as the primary goal in the past, and it regards achievements in GDP development as the promotion standard of local government officials [1]. Promotion incentives enable local governments to accelerate economic development through land management [2].

There are three mechanisms for local governments to develop the economy through land management. Firstly, local governments attract enterprises and residents to invest in urban construction through the land resources mismatch. As the local governments control the supply of the primary land market, the price of commercial residential land is much higher than that of industrial land [3–5]. The low price

of industrial land has attracted enterprise investment, while the increasing price of commercial and residential land has attracted residents to invest in real estate; these investments boost economic growth [6, 7]. However, local governments attract investment through low-price competition, and the cheap industrial land lowers the entry threshold for companies, resulting in excess capacity and environmental pollution [8–10].

Secondly, local governments have gained more revenue through the price difference between industrial land and commercial residential land to promote economic growth, which is called land finance [11–13]. In the long-term process of developing the economy by operating land, the local government has become dependent on land finance and increased its income by continuously hoarding land and raising the price of land [7, 14]. This approach has created a speculative bubble in the real estate market, which is not sustainable. At the same time, land finance has also caused

many unsustainable problems such as environmental pollution and increased carbon emissions [15–17].

Thirdly, land development led by local governments accelerates urbanization and increases the rate of economic growth [18–20]. Land development expands urban areas and promotes economic growth through capital accumulation [21–23]. Urbanization concentrates population density and increases the ecological footprint, pushing the environment beyond carrying capacity [24–26]. In the process of industrialization, urbanization, and economic development, urban carbon emissions have increased significantly [27–31]. Some local governments blindly promote urbanization to develop the economy, resulting in many ghost towns [32]. The excessive urban expansion will also occupy many arable land resources, thus threatening the ecological environment and sustainable development [23, 33–35].

In general, urbanization, economic development, and land management led by the government will bring environmental problems, which have seriously hindered urban sustainable development. However, sustainable development involves more than just the environmental dimension; the Sustainable Development Goals system (SDGs) projected by the United Nations covers seventeen areas such as good health, quality education, and economic growth. As urbanization, land resources mismatch, and economic development bring more government revenue, China's sustainable development elements such as infrastructure, living standards, and public services have greatly improved [36–40]. Therefore, the urbanization led by the government will affect the level of urban sustainable development, and the direction of urban sustainable development depends on government expenditure preferences and land development planning. If local governments intend to channel revenues more to growth-oriented infrastructure, it will crowd out expenditure on welfare programs and public services [41]. Infrastructure investment will further increase government revenue and form internal strengthening mechanisms [42, 43], which cause excessive urbanization and hinder urban sustainable development. If the local government provides more land for the construction of public services and intends to channel revenues more to industrial restructuring, the living conditions of residents will be improved, and environmental pollution and carbon emissions will be reduced [44, 45].

The contributions of this paper can be summarized as follows: (1) The complete information dynamic game model we build shows a new theory about government revenue expansion, land resources mismatch, industrial structure adjustment, and urban sustainable development. (2) The random forest model was selected to fit the nonlinear relationship between government revenue expansion, land resources mismatch, industrial structure adjustment, and urban sustainable development. (3) More detailed conclusions are drawn from the study of the eastern, central, and western representative urban agglomerations.

In the rest of this paper, the second section introduces the theoretical model; the third section introduces methodology and data; the fourth section shows the results and discussion; the fifth section gives the conclusions.

Table 1: A simple Nash equilibrium game of local government competition.

Local governments a/b	GDP development	Sustainable development
GDP development	$\rho_1 + \sigma_1, \rho_1 + \sigma_1$ [1]	$\rho_1 + \sigma_1, \rho_2 + \sigma_2$
Sustainable development	$\rho_1 + \sigma_1, \rho_2 + \sigma_2$	$\rho_2 + \sigma_2, \rho_2 + \sigma_2$

ρ_1 and σ_1 corresponding to economic growth mode. ρ_2 and σ_2 corresponding to sustainable development model.

2. Theoretical Model

The promotion incentive mechanism makes local governments compete in revenue expansion, urbanization, land management, and industrial development. This competition can be described as a simple Nash equilibrium game (Table 1). Suppose that local government A and local government B have two strategies of GDP development and sustainable development. Both strategies can make local governments obtain the payoff of regional GDP level improvement ρ and government revenue increases σ. The local government will develop strategies based on the total payoff $\rho + \sigma$. Under the economic development strategy, local governments channel revenues more to growth-oriented infrastructure, which promotes GDP growth, accelerates the process of urbanization, and brings additional revenue. As the sustainable development strategy requires the government to invest revenue in public services and industrial upgrading, this will slow down urbanization and economic growth and cause ρ, $\sigma_1 > \sigma_2$. Therefore, local governments that adopt the GDP development strategy can always obtain more payoffs. In the equilibrium state, local governments A and B will adopt the GDP development strategy, which maximizes the payoff for both sides.

However, the development of GDP with the realization of sustainable economic development is not achieved overnight, and industrial upgrading is a long process. The above Nash equilibrium cannot accurately reflect the game relationship of local governments. Therefore, this study introduces the complete information dynamic game model to demonstrate the game relationship between local governments [46].

On the premise of completely taking economic development as the goal, the utility function of local government can be expressed as

$$U_i = \beta x_i + \theta_1 y_i - z_i. \tag{1}$$

U_i is the total utility of local government. x_i represents the revenue scale of local governments (including general public budget revenue, social insurance revenue, and land revenue). Urbanization will promote economic development and increase government revenue. β is the marginal utility of government revenue, including public services, infrastructure investment to stimulate economic growth, making up the deficit, and official promotion. y_i is GDP. θ_1 is the marginal utility of GDP development. z_i is the negative impact of economic development and urbanization, such as arable land degradation, environmental pollution, and overcapacity.

Since the central government has put forward the goal of sustainable development, we assume that local governments choose industrial upgrading with a ratio α. θ_2 is the marginal utility of industrial upgrading. The size of θ_1 and θ_2 depends on the development strategy of the central government. The utility function of local government can be expressed as

$$U_i = \beta(1-\alpha)x_i + \theta_1(1-\alpha)y_i - z_i + \theta_2\alpha y_i. \tag{2}$$

Industrial upgrading needs government support, so the government revenue is reduced to $(1-\alpha)x_i$. The competition between local governments will produce negative spillover effects, and the actual utility of local governments 1 and 2 can be expressed as

$$U_1' = U_1 - \gamma_1 U_2, \tag{3}$$

$$U_2' = U_2 - \gamma_2 U_1. \tag{4}$$

The relationship between U_1' and α_1 can be expressed as

$$\frac{dU_1'}{d\alpha_1} = -\beta x_1 - (\theta_1 - \theta_2)y_1. \tag{5}$$

When $\theta_1 > \theta_2$ and $dU_1'/d\alpha_1 < 0$, reducing the industrial upgrading ratio α_1 will increase the actual utility of local government 1. The relationship between U_1' and α_2 can be expressed as

$$\frac{dU_1'}{d\alpha_2} = \gamma_1 [\beta x_2 + (\theta_1 - \theta_2)y_2]. \tag{6}$$

When $\theta_1 > \theta_2\theta_1 > \theta_2$ and $(dU_1'/d\alpha_2) > 0$, reducing industrial upgrading ratio α_2 will reduce the actual utility of local government 1. Equations (5) and (6) show that if the local government pays more attention to GDP development, reducing the industrial upgrading ratio will increase its own actual utility and reduce the actual utility of other local governments. This will lead local governments to abandon the industrial upgrading strategy and join the GDP development competition.

However, as the local government revenue rises rapidly, the central government will penalize this behavior to protect the economy, and it leads to $\beta < 0$. If the penalization is sufficiently severe, equations (5) and (6) will be less than 0. Local government 1 will increase the industrial upgrading ratio α_1 to improve its own actual utility, and it will reduce the actual utility of local government 2.

Government expenditure cannot permanently stimulate the GDP development, and urbanization also has bottlenecks, so we assume that x_i and y_i have the following function:

$$y_i = c \ln x_i. \tag{7}$$

Studies have proved that China's carbon emissions, environment, and other sustainable development factors have Kuznets Curve [47–49]. We assume that z_i and y_i have the following function:

$$z_i = ay_i - by_i^2. \tag{8}$$

a, b, and c are positive parameters. Taking equations (7) and (8) into equation (6),

$$U_i = \beta(1-\alpha)x_i + \theta_1(1-\alpha)c\ln x_i$$
$$- \left(a(1-\alpha)c\ln x_i - b(1-\alpha)^2 c^2 (\ln x_i)^2\right) + \theta_2\alpha c\ln x_i. \tag{9}$$

The first-order condition can be obtained by taking the derivative of x_i:

$$\frac{dU_i}{dx_i} = \beta(1-\alpha) + [(\theta_1 - \theta_2 - a)(1-\alpha) + \theta_2$$
$$+ 2b(1-\alpha)^2 c \ln x_i]\frac{c}{x_i}. \tag{10}$$

When $\beta > 0$, if $\theta_1 - \theta_2 > a$, $(dU_i/dU_i) > 0$. It shows that when the marginal utility of GDP development is bigger than that of industrial upgrading, the local government will improve its own utility by expanding the scale of government revenue. However, as x_i increases large enough, the second term on the right of equation (10) approaches 0, and the symbol dU_i/dU_i depends on the financial supervision strategy of the central government β. If the central government takes punitive measures ($\beta < 0$), local governments will decrease their revenue. If the central government chooses to ignore ($\beta > 0$), local governments will continue to increase the scale of revenue.

Equation (10) can be transformed as

$$\frac{dU_i}{dx_i} = (1-\alpha)\left[\beta + 2b(1-\alpha)\frac{c^2}{x_i}\ln x_i\right]$$
$$+ [\theta_1(1-\alpha) + \alpha\theta_2 - a(1-\alpha)]\frac{c}{x_i}. \tag{11}$$

If $\forall x_i$, there is $dU_i/dx_i < 0$, then β, θ_1, and θ_2 need to satisfy two conditions. Firstly, $\beta < - [2b(1-\alpha)c^2\ln x_i]/x_i\beta < - [2b(1-\alpha)c^2\ln x_i]/x_i$; it means that the marginal utility of government revenue is negative, and the central government has strengthened financial supervision. Secondly, $\theta_1 < - [\alpha/(1-\alpha)\theta_2]\theta_1 < - [\alpha/(1-\alpha)\theta_2]$; it means that the marginal utility of GDP government is negative, and local government will choose the industrial upgrading strategy. These two conditions are indispensable. If the GDP development encounters a bottleneck and the central government chooses a loose financial supervision strategy, local governments will still maintain GDP growth by expanding the scale of government revenue. This will lead to unsustainable problems such as land resources mismatch and overcapacity; if the GDP development lags behind and the central government chooses a strict financial supervision strategy, the development enthusiasm of local governments will be reduced. Insufficient fiscal revenue will prevent local governments from providing adequate public services such as healthcare and education.

For the central government, the increase in revenue can improve the level of public facilities such as national defense, infrastructure, medical care, and education and promote economic growth to improve people's living standards. For local governments, the increase in fiscal revenue can achieve

TABLE 2: Composition indicators of USDI.

Indicator	Unit	Connotation	Effects
GDP per capita	RMB	Economic development level	Positive
Green coverage	%	Urban greening level	Positive
Carbon emissions per capita	Ton/p	Energy consumption	Negative
Industrial waste gas emissions per capita	Ton/10000 p	Air pollution	Negative
Industrial wastewater discharge per capita	Ton/p	Water pollution	Negative
Proportion of primary and secondary school teachers	‰	Education	Positive
Proportion of doctors	‰	Medical and health	Positive

the GDP target set by the central government and reduce the deficit. However, development aimed at increasing fiscal revenue has drawbacks. In pursuit of the dual goals of GDP scale and government revenue scale, high pollution and high energy consumption have appeared, and this phenomenon has entered a vicious circle.

In general, the financial supervision strategy and economic development strategy of the central government will affect the decision-making of local governments in industrial structure adjustment and revenue adjustment. Different development stages should correspond to different strategies. In the early stage of economic development, the central government adopts a loose fiscal supervision strategy and a GDP development strategy, and local governments can increase government revenue through the land resources mismatch to develop GDP and provide more public services. As the GDP development encounters bottlenecks, the central government should adopt a strict financial supervision strategy and a sustainable development strategy, and local governments should reduce government revenue, adjust industrial structure, and solve the problem of land resources mismatch.

As China's GDP growth rate declines, the government needs to actively change the way of economic development. Local governments should not rely on developing GDP by land management, especially for the land resources mismatch. Researches show that land transfer marketization can improve the industrial structure and reduce environmental pollution [50–52]. Therefore, local governments should make efforts to control the scale of government revenue, reduce the investment in growth-oriented infrastructure, provide more public services, and support industrial upgrading.

3. Methodology and Data

3.1. Urban Sustainable Development Index (USDI)

3.1.1. Indicators Selection. Studies on sustainability measurement are mainly based on provincial data [53–55]. Based on the perspective of land resources mismatch, the competition between local governments is more obvious. Therefore, this study attempts to measure the level of urban sustainable development. However, multidimensional evaluation can accurately calculate the level of urban sustainable development. However, due to the limited integrity of urban data, we comprehensively consider the integrity and diversity of the data. The basis of data selection is as

follows: Firstly, refer to the SDGs sustainable index evaluation system. Secondly, select widely used indicators. Thirdly, select representative indicators. Finally, we have selected 7 indicators to form an urban sustainable development index (Table 2).

The indicators cover three aspects of economy, society, and environment. All indicators include 292 city data from 2000 to 2019. The carbon emissions data were obtained from the research of Chen [56]. Other data were obtained from the China Urban Statistical Yearbook. The missing data are supplemented by the interpolation method.

3.1.2. Calculation. All indicators need to be normalized to eliminate dimensional discrepancies, and the normalization equations are as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{12}$$

$$x' = \frac{x_{\max} - x}{x_{\max} - x_{\min}}. \tag{13}$$

Equations (12) and (13) apply to positive and negative indicators, respectively; the standardized indicators are within the interval [0, 1]. However, the geometric mean method with a standardized indicator of zero would result in an SDI of zero. We need to set a minimum value to ensure that all standardized indicators are within the interval (0, 1]. The upper limit also needs to be set for indicators of sustainable development.

The USDI is established using the geometric mean method with reference to the method of the Human Sustainable Development Index (HSDI) [57, 58]. The calculation formula is as follows:

$$\text{USDI} = \sqrt[7]{\prod_{i=1}^{7} x_i}. \tag{14}$$

x_i is the value of the indicator after normalization. The reason for using the geometric mean calculation is that there should be a nonlinear substitution relationship between indicators, and the geometric mean method can penalize cities with unbalanced development [59–61].

3.2. Selection of Explanatory Variables. Explanatory variables include industrial structure (IS), government revenue level GR(GR), and degree of land resources mismatch (LRM). The

proportion of the tertiary industry GDP in different cities in China varies greatly, so the proportion of GDP of the tertiary industry is used to represent the industrial structure.

The government revenue level is represented by the ratio of local government revenue to GDP. Local government revenue includes local governments' general public budget revenue (GPBR), social insurance funds revenue (SIFR), and land revenue (LR).

Land transfer methods in China include agreement, tenders, auction, and listing transfers. Tenders, auctions, and listings for sale can get higher unit prices due to the market-oriented transfer. Therefore, the government will use the lower-priced agreement land to attract investment and then transfer the higher-priced tender auction and listing land to commercial and residential real estate. This way of selling is called land resources mismatch. Firstly, it can develop GDP by attracting investment; secondly, it can increase the local government revenue by raising the market price of land. In summary, the degree of land resources mismatch can be expressed as follows:

$$\mathrm{LRM}_{it} = \begin{cases} \dfrac{\mathrm{AGR}_{it}}{\mathrm{TAL}_{it}}, & \mathrm{LRM}_{it} \leq 1, \\ \\ 1, & \mathrm{LRM}_{it} > 1. \end{cases} \quad (15)$$

AGR_{it} is the unit prices of agreement transfer, and TAL_{it} is the unit prices of tender, auction, and listing transfer. As there is a correlation between annual land revenue and the unit price gap, we believe that the unit price ratio is more suitable for measuring the degree of land resource mismatch than the area ratio, which is different from existing research methods [10, 62]. Normally, land buyers refuse to bear the premium, but due to the small scale of land transfer in some cities, there are special agreements for sales in some years that cause $\mathrm{AGR}_{it} > \mathrm{TAL}_{it}$. Therefore, when $\mathrm{LRM}_{it} > 1$, it means that there is no land resources mismatch.

Since there are many factors affecting the level of sustainable development, the absence of variables will cause regression bias. Therefore, auxiliary variables need to be added to improve the goodness of fit of the model, which can improve the prediction accuracy of the model. We attempted to add population density (PD), geographic location (GL), real estate development level (RED), and average output of industrial enterprises (AOIE) as auxiliary explanatory variables. All these factors will affect the government's decision-making in urban sustainable construction. For example, the local government will plan the number of hospitals and schools based on the city's population density; the local government will plan the proportion of residential, commercial, and industrial land for the next year based on the real estate development situation; the local government will decide whether to close polluting enterprises based on the urban environment. However, adding auxiliary explanatory variables does not necessarily improve the accuracy of the model because the added variables may not be relevant to urban sustainability. We can select the variables based on which combination performs better. If the added variables did not improve the goodness of fit of the model, we chose to ignore these variables.

IS, GPBR, PD, RED, and AOIE data are from China Urban Statistical Yearbook. SIFR data are from China Statistical Yearbook. LR, AGR, and TAL data are from China Land Resources Statistical Yearbook.

3.3. Machine Learning Method-Random Forest.

The random forest model is a kind of machine learning method proposed by Breiman [63]. It consists of multiple decision trees, and each decision tree is independent (Figure 1).

The random forest model is an ensemble learning and bootstrap sampling process; suppose the original sample set contains N samples and M features. Firstly, for each decision tree, repeatedly and randomly select N samples from the original sample set with replacements to generate a new training sample set. Repeated sampling will cause some samples not to participate in the splitting of the decision tree, which is called out of bag (OOB). OOB is used to estimate the misclassification rate of decision trees. For each node of the decision tree, randomly select m features as the split basis ($m \ll M$). Secondly, using the Classification and Regression Tree (CART) with algorithm regressing each tree, the objective function is as follows:

$$\min \sum_{j=1}^{J} \sum_{i \in R_j} \left(y_i - \widehat{y}_{R_j} \right)^2. \quad (16)$$

R_j represents dividing the sample into j distinct regions, and $\widehat{y}_{R_j}$ represents the average predicted value in R_j. Different from the traditional CART algorithm, the random forest model requires the decision trees split completely and ensure the integrity of the trees. Thirdly, take the average of all decision tree predictions as the prediction result of the random forest model. Model prediction accuracy is evaluated by root mean squared error (RMSE) and goodness of fit (R^2).

The reasons for choosing the random forest model in this study were as follows: (1) Studies have shown that the relationship between urban development, land resource allocation, and sustainability factors such as air quality, carbon emissions, and environmental pollution is nonlinear. Due to the antioverfitting and antinoise capabilities, the random forest model is widely used in the study of nonlinear problems [64–70]. (2) The construction decision-making process of local governments in urbanization is similar to decision trees in random forests. For example, when land resources are expropriated, how many hospitals should be planned? Hundreds of local governments will make decisions based on urban conditions (Figure 2), and the mean value of local government plans in similar cities is the predicted value.

To improve the accuracy of the prediction results, we use three models for random forest regression. Model 1 is the base model, consisting of USDI, GR, LRM, and IS:

$$\mathrm{USDI} \sim \mathrm{GR} + \mathrm{LRM} + \mathrm{IS}. \quad (17)$$

Model 2 adds PD, GL, RED, and AOIE, supplementing other factors affecting sustainable development in the process of urbanization:
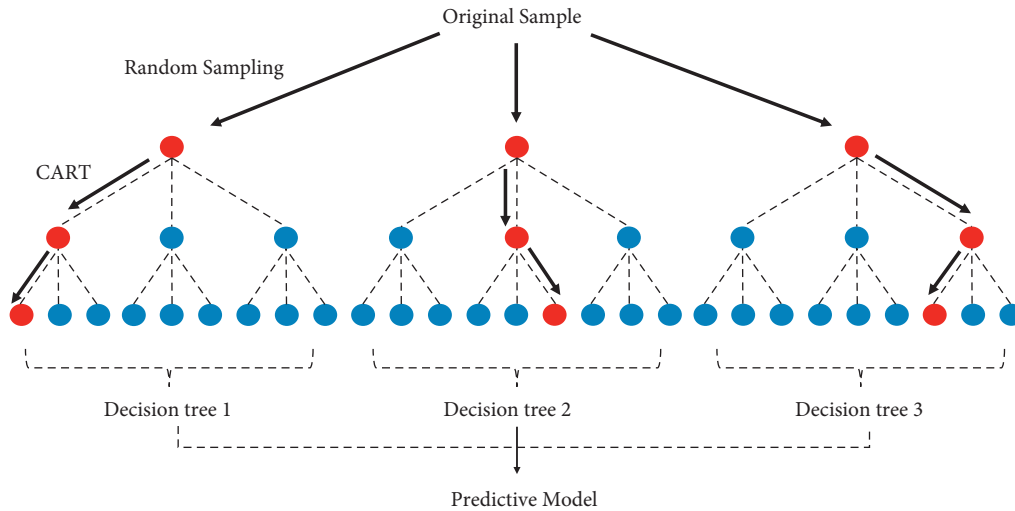
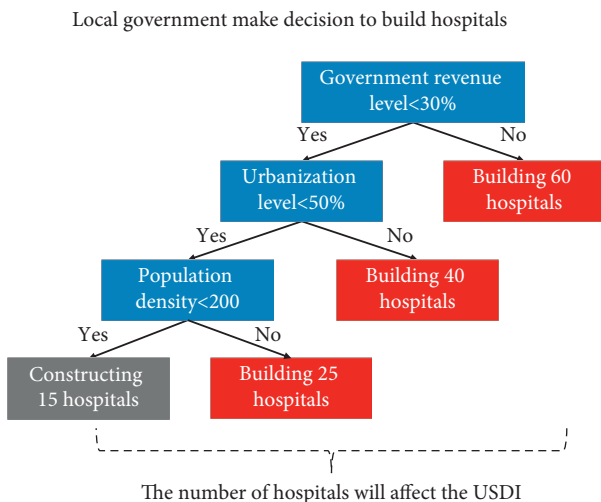FIGURE 1: The principle of the random forest model.

FIGURE 2: An example of a local government decision affecting sustainable development.

$$USDI \sim GR + LRM + IS + PD + GL + RED + AOIE. \quad (18)$$

Model 3 splits the GR to study the impact of different revenues on USDI:

$$USDI \sim GPBR + SIFR + LR + LRM + IS + PD + GL + RED + AOIE. \quad (19)$$

Extreme gradient boosting model (XGBoost) and multiple linear regression (MLR) will be used to test the robustness of all models.

4. Results and Discussion

4.1. Predictive Performance of Random Forest Model. We divide the original sample into two parts: 70% of the data are used for training, and 30% of the data are used for testing. The training set is used for model regression, and the testing set is used to examine the generalization ability of the model.

Table 3 shows that R^2 of the base model on the test set is only 45.9%. By adding auxiliary explanatory variables, R^2 of the test set can be improved to 70.4%. Further splitting the GR improves R^2 of the test set to 74.0%. Compared with the split of GR, the auxiliary explanatory variables have a greater improvement in the fit of the prediction set. The addition of auxiliary explanatory variables significantly improves the generalization ability of the model and reduces the dispersion of the prediction set (Figure 3). The robustness test (Table 3) shows that the performance of the random forest model is better than that of XGboost or MLR.

Figure 4 shows the importance of variables in a different model. In the base model, GR has the highest importance. After adding auxiliary explanatory variables, IS becomes the most important variable in model, indicating that the synergy of AOIE and IS has a higher impact on sustainable development than GR. The increase of AOIE and IS will reduce the discharge of industrial wastewater and exhaust gas, which will affect the USDI. After the split of GR, the importance of LRM has increased, probably because of the synergistic effect with LR. Local governments will increase the LR by changing the LRM strategy. The importance of AOIE decreases, while the importance of SIFR is similar to that of AOIE. SIFR is highly correlated with the number of hospitals and schools, and AOIE is highly correlated with industrial wastewater and air emissions, so both have the same impact on USDI. PD and GL both have great importance to USDI, which has not changed before or after splitting GR. Urbanization will increase the number and density of urban populations, and local governments will improve healthcare and education. China's eastern cities have more permanent residents than registered households, and these cities have higher levels of medical care and education (the medical level and education level variables use the registered population). Therefore, both geographic location and population density of cities will affect USDI. The importance of IS still ranks first, indicating that the adjustment of industrial structure is the most important for the urban sustainable development.

TABLE 3: Robustness test of random forest model.

Model	Random forest				XGBoost				MLR
	R^2		RMSE*		R^2		RMSE		R^2
	Train	Test	Train	Test	Train	Test	Train	Test	
1	0.913	0.459	0.013	0.030	0.526	0.504	0.027	0.027	0.427
2	0.957	0.704	0.009	0.022	0.540	0.474	0.027	0.029	0.506
3	0.958	0.740	0.009	0.020	0.794	0.519	0.018	0.027	0.508

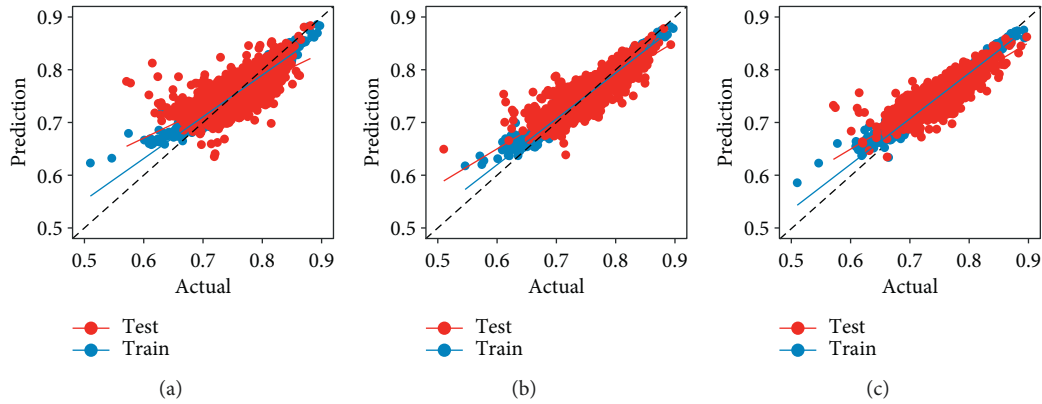*RMSE is the root mean square error.

FIGURE 3: The generalization ability of the three models. (a) Representative model 1, (b) representative model 2, and (c) representative model 3.
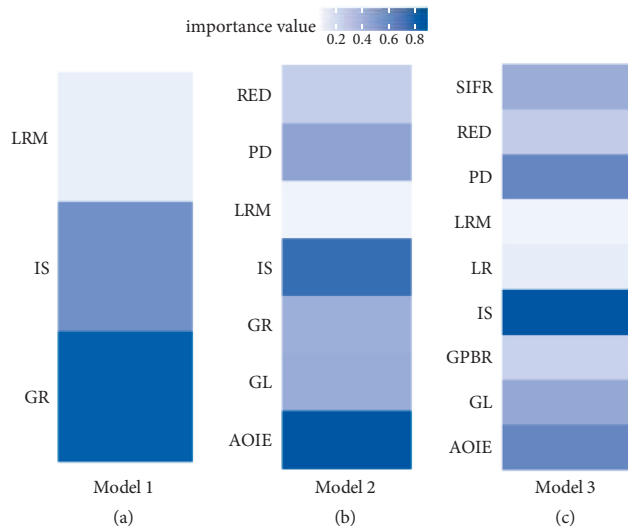

FIGURE 4: The importance of variables of the three models. (a) Model 1, (b) model 2, and (c) model 3.

heterogeneity. The importance value of the random forest model is obtained by adding noise interference to the OOB. Therefore, the smaller the importance value of the feature, the stronger the noise of the dataset. The strong noise of the variable indicates that the data tend to be randomly distributed in different years. For policy variables, the strong noise indicates that local governments have large changes in policy designation each year. For example, the importance of LRM in the mid is the highest, indicating that the LRM strategies of the midlocal governments are regular. In contrast, the LRM strategies of eastern and western local governments have great volatility. However, due to the large difference in land prices between eastern and western regions, the LRM strategies of eastern and western local governments should be different. The eastern local government maximizes LR by continuously adjusting the LRM strategy, while the western local government cannot misallocate land resources.

AOIE, IS, and PD show great importance in both models, which verifies the conclusion in 4.1. Regarding the analysis of the theoretical mechanism, PD determines the level of education and medical care in the city, IS determines the level of cities' economic development, and AOIE determines the level of wastewater and waste pollution in the city. The local government should focus on improving these three indicators to increase USDI, for example, reducing the number of agreements and industrial land transfers and increasing the AOIE; actively adjusting the industrial structure and improving the level of local economic development; rationally increasing the construction of hospitals and schools in the process of urbanization.

4.2. Regional Heterogeneity Research. Dividing the data into east, middle, and west for random forest regression further improves the generalization ability of the model, and the performance of the random forest algorithm is better than that of XGboost or MLR (Table 4). Since the random forest model performs better in models 2 and 3, the following studies will exclude model 1.

In model 2, the importance of PD, GR, RED, and LRM in the eastern, mid, and western regions makes heterogeneity (Figure 5). After splitting GR, LRM still shows great

TABLE 4: Robustness test of random forest model.

| Model	Location	Random forest				XGBoost				MLR
		R^2		RMSE		R^2		RMSE		R^2
		Train	Test	Train	Test	Train	Test	Train	Test	
2	East	0.966	0.803	0.007	0.017	0.893	0.792	0.012	0.018	0.667
	West	0.948	0.690	0.010	0.022	0.779	0.648	0.019	0.023	0.480
	Mid	0.957	0.754	0.008	0.018	0.846	0.682	0.014	0.020	0.467
3	East	0.970	0.825	0.007	0.016	0.902	0.796	0.012	0.018	0.668
	West	0.956	0.759	0.011	0.024	0.900	0.731	0.016	0.022	0.413
	Mid	0.959	0.684	0.008	0.020	0.900	0.680	0.011	0.020	0.442
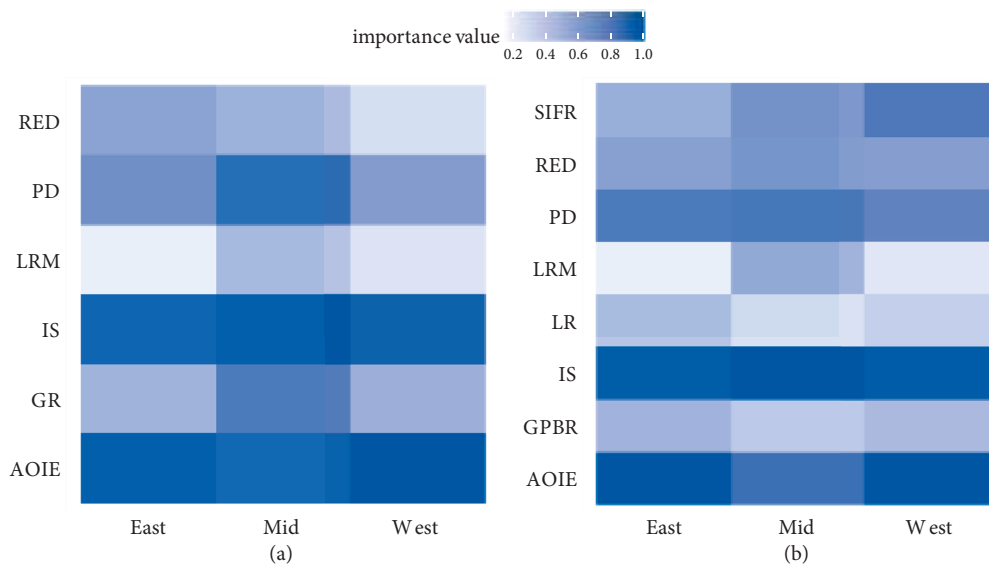
FIGURE 5: The importance of variables of the three models. (a) Corresponding to model 2; (b) corresponding to model 3.

4.3. Representative Urban Agglomeration Research.
Beijing-Tianjin-Hebei (BTH), Yangtze River Delta (YRD), Pearl River Delta (PRD), and Chengdu-Chongqing (CC) are the four most representative urban agglomerations in China. These four urban agglomerations account for 36.3% of China's population and 47.3% of China's GDP. Table 5 shows that the generalization ability of the random forest model is improved by choosing the representative urban agglomerations. The performance of the random forest model is still better than XGBoost and MLR.

Figure 6 shows that IS and AOIE have a strong impact on the USDI of the four urban agglomerations, which strengthens the previous results. The GR of these urban agglomerations has a strong impact on the USDI. Due to the strong ability of land resource control, the local governments will adjust LR and LRM strategies to cover excess spending. Therefore, the impact of LRM and LR in these urban agglomerations on USDI is not obvious, but the GR is contrary. The importance of SIFR in BTH and CCCC is obvious, which is different from YRD and PRD. The RED and PD of the four urban circles are stable in both models.

4.4. Discussion of Results.
From the regression results, IS has the greatest impact on urban sustainable development. Because cities with a high proportion of tertiary industry mean a small proportion of industrial output, there will be fewer emissions of waste gas and wastewater. Therefore, the government should actively adjust the industrial structure to improve the level of sustainable urban development. However, the industry is the foundation of national economic development. Excessively increasing proportion of the tertiary industry will lead to the relocation and loss of manufacturing. Therefore, for cities with a high proportion of industrial output, local governments should focus on improving the output quality and efficiency of industrial enterprises to improve the level of urban sustainable development. The importance of AOIE to urban sustainable development confirms it. Local governments should not continue to increase the number of industrial enterprises through the land resources mismatch but increase output and reduce pollution by providing tax incentives and R&D subsidies for enterprises. Surprisingly, SIFR showed a great impact in BTH and CC and was as important in PRD and YRD as AOIE. Local governments need to work hard to improve the level of urban sustainable development by improving the level of SIFR. However, raising the SIFR level will increase the tax burden for both businesses and residents.

TABLE 5: Robustness test of random forest model.

| Model	Location	Random forest				XGBoost				MLR
		R^2		RMSE		R^2		RMSE		R^2
		Train	Test	Train	Test	Train	Test	Train	Test	
2	BTH	0.982	0.896	0.005	0.011	0.951	0.879	0.009	0.011	0.838
	YRD	0.971	0.839	0.006	0.015	0.902	0.810	0.011	0.017	0.676
	PRD	0.980	0.770	0.006	0.018	0.920	0.903	0.011	0.012	0.767
	CC	0.974	0.847	0.007	0.017	0.924	0.846	0.011	0.017	0.632
3	BTH	0.989	0.913	0.004	0.012	0.953	0.911	0.008	0.010	0.903
	YRD	0.967	0.736	0.008	0.018	0.895	0.766	0.012	0.018	0.555
	PRD	0.979	0.889	0.006	0.014	0.949	0.854	0.009	0.015	0.801
	CC	0.975	0.879	0.007	0.014	0.901	0.827	0.013	0.017	0.733
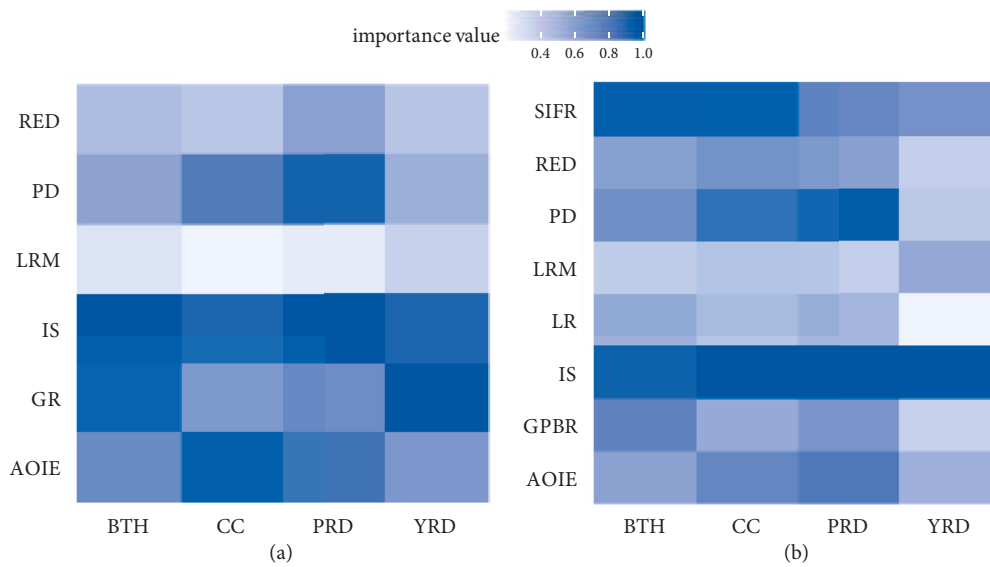
FIGURE 6: The importance of variables of the three models. (a) Corresponding to model 2; (b) corresponding to model 3.

Local governments should find a compromise solution to this problem.

The impact of LRM and LR on urban sustainable development is not as important as expected, probably because the intervention of local governments in land supply makes land prices cyclical. The purpose of local government intervention in land supply is to maximize government revenue, which is confirmed by the importance of GR. The importance of LRM and LR gradually increases as the region shrinks. This shows that the local government's land resource allocation strategy is affected by the neighboring cities.

5. Conclusions

This paper discusses the impact of government revenue expansion, land resource mismatch, and industrial structure adjustment on sustainable urban development in the process of urbanization. Our research shows that the adjustment of industrial structure and the level of local government income have a great impact on urban sustainable development. The importance of land resources mismatch is stronger in representative metropolitan research. The author believes that land resources mismatch is a measure adopted by local governments to increase the level of income, which is highly cyclical and difficult to detect. The government's influence on sustainable development through land management is reflected in the government revenue.

Urbanization led by local governments will affect sustainable development in many aspects. Local governments will plan land construction after land requisition, including planning the area of industrial, commercial, and residential land, the number of schools and hospitals, public greening, and roads. Land construction planning has different impacts on urban sustainable development. For example, industrial enterprises will cause environmental pollution due to the discharge of waste gas and wastewater; population aggregation will improve the level of medical care and education. Therefore, how local governments plan for land determines the direction of sustainable urban development. Since the reduction in land resources mismatch will reduce the local government revenue, the local government will choose to reduce the expenditure on public services to maintain the speed of economic development. Therefore, the central government should change the promotion incentives and encourage local governments to devote more resources to

urban sustainable development. And the central government needs to undertake the corresponding sustainable development construction tasks instead of handing all the tasks to the local governments. This is because the local government adopts a fiscal expansion strategy due to the excess construction tasks. After the central government's incentive mechanism changed, local governments should formulate development strategies according to the industrial structure of the city, not just maximizing the speed of economic development and maximizing government revenue. In the process of industrial transformation and upgrading, the normal operation of the urban economy must be ensured first, and blind and radical strategies should not be adopted.

In addition, local governments should be committed to improving the level of social insurance income in cities, but they should not increase the tax burden on businesses and residents. We think the best way is to let state-owned enterprises take this responsibility. Due to the government's endorsement, state-owned enterprises have advantages over private enterprises in financing and procurement. Therefore, state-owned enterprises are equivalent to enjoying the preferential treatment of market resource allocation. The specific plan is to transfer part of the annual profits of state-owned enterprises to the social insurance fund.

Our study shows that random forest models perform better in government decision-making problems, but how to select variables to measure the level of urban sustainable development more accurately deserves further research. In future research, we will further improve the urban sustainable development index system, and the measurement of land resource mismatch will also adopt a variety of methods.

Data Availability

All data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares no conflicts of interest.

Authors' Contributions

Yixiao Peng contributed to conceptualization, methodology, software, formal analysis, data curation, original draft preparation, and visualization.

References

[1] Y. Shi, C. P. Chang, C. L. Jang, C. L. Jiang, and Y. Hao, "Does economic performance affect officials' turnover? evidence from municipal government leaders in China," *Quality and Quantity*, vol. 52, no. 4, pp. 1873–1891, 2018.

[2] Z. Chen, J. Tang, J. Wan, and Y. Chen, "Promotion incentives for local officials and the expansion of urban construction land in China: using the Yangtze river Delta as a case study," *Land Use Policy*, vol. 63, pp. 214–225, 2017.

[3] L. Tian and W. Ma, "Government intervention in city development of China: a tool of land supply," *Land Use Policy*, vol. 26, no. 3, pp. 599–609, 2009.

[4] T. Zhong, Y. Chen, and X. Huang, "Impact of land revenue on the urban land growth toward decreasing population density in Jiangsu province, China," *Habitat International*, vol. 58, pp. 34–41, 2016.

[5] J. Zhou, X. Yu, X. Jin, and N. Mao, "Government competition, land supply structure and semi-urbanization in China," *Land*, vol. 10, no. 12, p. 1371, 2021.

[6] C. He, Z. Huang, and R. Wang, "Land use change and economic growth in urban China: a structural equation analysis," *Urban Studies*, vol. 51, no. 13, pp. 2880–2898, 2014.

[7] J. Zhang, T. Yu, L. Li, D. Zhang, G. Zhao, and H. Wen, "Marketization allocation, land price, and local government land speculation, China. Int," *International Journal of Strategic Property Management*, vol. 24, no. 5, pp. 335–347, 2020.

[8] L. Tian, "Land use dynamics driven by rural industrialization and land finance in the peri-urban areas of China: "The examples of Jiangyin and Shunde"," *Land Use Policy*, vol. 45, pp. 117–127, 2015.

[9] Z. Huang and X. Du, "Strategic interaction in local governments' industrial land supply: evidence from China," *Urban Studies*, vol. 54, no. 6, pp. 1328–1346, 2017.

[10] W. Du and M. Li, "The impact of land resource mismatch and land marketization on pollution emissions of industrial enterprises in China," *Journal of Environmental Management*, vol. 299, Article ID 113565, 2021.

[11] G. C. S. Lin and F. Yi, "Urbanization of capital or capitalization on urban land? land development and local public finance in urbanizing China," *Urban Geography*, vol. 32, no. 1, pp. 50–79, 2011.

[12] W. Qun, L. Yongle, and Y. Siqi, "The incentives of China's urban land finance," *Land Use Policy*, vol. 42, pp. 432–442, 2015.

[13] Y. Ji, X. Guo, S. Zhong, and L. Wu, "Land financialization, uncoordinated development of population urbanization and land urbanization, and economic growth: evidence from China," *Land*, vol. 9, no. 12, p. 481, 2020.

[14] J. Du and R. B. Peiser, "Land supply, pricing and local governments' land hoarding in China," *Regional Science and Urban Economics*, vol. 48, pp. 180–189, 2014.

[15] W. Zhang and H. Xu, "Effects of land urbanization and land finance on carbon emissions: a panel data analysis for Chinese provinces," *Land Use Policy*, vol. 63, pp. 493–500, 2017.

[16] L. O. Wang, H. Wu, and Y. Hao, "How does China's land finance affect its carbon emissions?" *Structural Change and Economic Dynamics*, vol. 54, pp. 267–281, 2020.

[17] J. Wang, F. Gao, P. Tang, E. van der Krabben, H. Ploegmakers, and A. Samsura, "Will the increase in local land-based revenues lead to an improvement in environmental quality? an empirical study through the lens of land finance in China," *Journal of Environmental Planning and Management*, pp. 1–28, 2022.

[18] G. C. S. Lin, "China's landed urbanization: neoliberalizing politics, land commodification, and municipal finance in the growth of metropolises," *Environment & Planning A*, vol. 46, no. 8, pp. 1814–1835, 2014.

[19] Y. Liu, B. Yan, and Y. Zhou, "Urbanization, economic growth, and carbon dioxide emissions in China: a panel cointegration and causality analysis," *Journal of Geographical Sciences*, vol. 26, no. 2, pp. 131–152, 2016.

[20] F. Wu, "Land financialisation and the financing of urban development in China," *Land Use Policy*, vol. 112, Article ID 104412, 2022.

[21] X. Deng, J. Huang, S. Rozelle, and E. Uchida, "Economic growth and the expansion of urban land in China," *Urban Studies*, vol. 47, no. 4, pp. 813–843, 2010.

[22] L. Ye and A. M. Wu, "Urbanization, land development, and land financing: evidence from Chinese cities," *Journal of Urban Affairs*, vol. 36, no. 1, pp. 354–368, 2014.

[23] W. Liang and M. Yang, "Urbanization, economic growth and environmental pollution: evidence from China," *Sustainable Computing: Informatics and Systems*, vol. 21, pp. 1–9, 2019.

[24] J. Zhu, "The impact of land rent seeking and dissipation during institutional transition on China's urbanization," *Urban Affairs Review*, vol. 53, no. 4, pp. 689–717, 2017.

[25] S. T. Hassan, M. A. Baloch, N. Mahmood, J. Zhang, and J. Zhang, "Linking economic growth and ecological footprint through human capital and biocapacity," *Sustainable Cities and Society*, vol. 47, Article ID 101516, 2019.

[26] Z. Ahmed, M. M. Asghar, M. N. Malik, and K. Nawaz, "Moving towards a sustainable environment: the dynamic linkage between natural resources, human capital, urbanization, economic growth, and ecological footprint in China," *Resources Policy*, vol. 67, Article ID 101677, 2020.

[27] J. Ou, X. Liu, X. Li, and Y. Chen, "Quantifying the relationship between urban forms and carbon emissions using panel data analysis," *Landscape Ecology*, vol. 28, no. 10, pp. 1889–1907, 2013.

[28] S. Wang, C. Fang, H. Ma, Y. Wang, and J. Qin, "Spatial differences and multi-mechanism of carbon footprint based on GWR model in provincial China," *Journal of Geographical Sciences*, vol. 24, no. 4, pp. 612–630, 2014.

[29] B. Xu and B. Lin, "How industrialization and urbanization process impacts on CO_2 emissions in China: evidence from nonparametric additive regression models," *Energy Economics*, vol. 48, pp. 188–202, 2015.

[30] H. Xu and W. Zhang, "The causal relationship between carbon emissions and land urbanization quality: a panel data analysis for Chinese provinces," *Journal of Cleaner Production*, vol. 137, pp. 241–248, 2016.

[31] L. Li, X. Hong, D. Tang, and M. Na, "GHG emissions, economic growth and urbanization: a spatial approach," *Sustainability*, vol. 8, no. 5, 2016.

[32] H. Zheng, X. Wang, and S. Cao, "The land finance model jeopardizes China's sustainable development," *Habitat International*, vol. 44, pp. 130–136, 2014.

[33] S. Guo, Y. Wang, Y. Wang, M. Wang, P. He, and L. Feng, "Inequality and collaboration in north China urban agglomeration: evidence from embodied cultivated land in Jing-Jin-Ji's interregional trade," *Journal of Environmental Management*, vol. 275, Article ID 111050, 2020.

[34] X. H. Lu and S. G. Ke, "Evaluating the effectiveness of sustainable urban land use in China from the perspective of sustainable urbanization," *Habitat International*, vol. 77, pp. 90–98, 2018.

[35] J. Du, J. C. Thill, R. B. Peiser, and C. Feng, "Urban land market and land-use changes in post-reform China: a case study of Beijing," *Landscape and Urban Planning*, vol. 124, pp. 118–128, 2014.

[36] X. Guan, H. Wei, S. Lu, Q. Dai, and H. Su, "Assessment on the urbanization strategy in China: achievements, challenges and reflections," *Habitat International*, vol. 71, pp. 97–109, 2018.

[37] M. Chen, C. Ye, D. Lu, Y. Sui, and S. Guo, "Cognition and construction of the theoretical connotations of new urbanization with Chinese characteristics," *Journal of Geographical Sciences*, vol. 29, no. 10, pp. 1681–1698, 2019.

[38] X. Liu, W. Fang, H. Li, X. Han, and H. Xiao, "Is urbanization good for the health of middle-aged and elderly people in China?-based on CHARLS data," *Sustainability*, vol. 13, no. 9, p. 4996, 2021.

[39] J. Lu, B. Li, and H. Li, "The influence of land finance and public service supply on peri-urbanization: evidence from the counties in China," *Habitat International*, vol. 92, Article ID 102039, 2019.

[40] G. G. Liu, X. Wu, C. Peng, and A. Z. Fu, "Urbanization and health care in rural China," *Contemporary Economic Policy*, vol. 21, no. 1, pp. 11–24, 2003.

[41] S. Guo and Y. Shi, "Infrastructure investment in China: a model of local government choice under land financing," *Journal of Asian Economics*, vol. 56, pp. 24–35, 2018.

[42] P. Tang, X. Shi, J. Gao, S. Feng, and F. Qu, "Demystifying the key for intoxicating land finance in China: an empirical study through the lens of government expenditure," *Land Use Policy*, vol. 85, pp. 302–309, 2019.

[43] C. Ding, Y. Niu, and E. Lichtenberg, "Spending preferences of local officials with off-budget land revenues of Chinese cities," *China Economic Review*, vol. 31, pp. 265–276, 2014.

[44] Z. F. Mi, S. Y. Pan, H. Yu, and Y. M. Wei, "Potential impacts of industrial structure on energy consumption and CO_2 emission: a case study of Beijing," *Journal of Cleaner Production*, vol. 103, pp. 455–462, 2015.

[45] Q. Lin, L. Zhang, B. Qiu, Y. Zhao, and C. Wei, "Spatiotemporal analysis of land use patterns on carbon emissions in China," *Land*, vol. 10, no. 2, p. 141, 2021.

[46] J. Ma, "Modelling central-local fiscal relations in China," *China Economic Review*, vol. 6, no. 1, pp. 105–136, 1995.

[47] Y. Dongfeng, Y. Chengzhi, and L. Ying, "Urbanization and sustainability in China: an analysis based on the urbanization Kuznets-curve," *Planning Theory*, vol. 12, no. 4, pp. 391–405, 2013.

[48] A. Jalil and S. F. Mahmud, "Environment Kuznets curve for CO_2 emissions: a cointegration analysis for China," *Energy Policy*, vol. 37, no. 12, pp. 5167–5172, 2009.

[49] T. Song, T. Zheng, and L. Tong, "An empirical test of the environmental Kuznets curve in China: a panel cointegration approach," *China Economic Review*, vol. 19, no. 3, pp. 381–392, 2008.

[50] X. H. Lu, X. Jiang, and M. q. Gong, "How land transfer marketization influence on green total factor productivity from the approach of industrial structure? evidence from China," *Land Use Policy*, vol. 95, Article ID 104610, 2020.

[51] W. Sun, Z. Chen, and D. Wang, "Can land marketization help reduce industrial pollution?" *International Journal of Environmental Research and Public Health*, vol. 16, no. 12, 2019.

[52] D. Wang, C. Ren, and T. Zhou, "Understanding the impact of land finance on industrial structure change in China: insights from a spatial econometric analysis," *Land Use Policy*, vol. 103, Article ID 105323, 2021.

[53] Z. Xu, S. N. Chau, X. Chen et al., "Assessing progress towards sustainable development over space and time," *Nature*, vol. 577, no. 7788, pp. 74–78, 2020.

[54] Y. Wang, Y. Lu, G. He, C. Wang, J. Yuan, and X. Cao, "Spatial variability of sustainable development goals in China: a provincial level evaluation," *Environmental Development*, vol. 35, Article ID 100483, 2020.

[55] Z. Zhao, Y. Pan, J. Zhu, J. Wu, and R. Zhu, "The impact of urbanization on the delivery of public service–related SDGs in China," *Sustainable Cities and Society*, vol. 80, Article ID 103776, 2022.

[56] J. Chen, M. Gao, S. Cheng et al., "County-level CO_2 emissions and sequestration in China during 1997–2017," *Scientific Data*, vol. 7, 2020.

[57] C. Togtokh, "Time to stop celebrating the polluters," *Nature*, vol. 479, no. 7373, 2011.

[58] G. Bravo, "The human sustainable development index: new calculations and a first critical analysis," *Ecological Indicators*, vol. 37, pp. 145–150, 2014.

[59] C. Herrero, R. Martínez, and A. Villar, "Multidimensional social evaluation: an application to the measurement of human development," *Review of Income and Wealth*, vol. 56, no. 3, pp. 483–497, 2010.

[60] C. Herrero, R. Martínez, and A. Villar, "A newer human development index," *Journal of Human Development and Capabilities*, vol. 13, no. 2, pp. 247–268, 2012.

[61] E. Zambrano, "An axiomatization of the human development index," *Social Choice and Welfare*, vol. 42, no. 4, pp. 853–872, 2014.

[62] L. Li, P. Huang, and G. Ma, "Mismatch of land resources and productivity difference of industrial enterprises in China," *Management World*, vol. 32, pp. 86–96, 2016.

[63] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[64] M. Emami Javanmard and S. F. Ghaderi, "A hybrid model with applying machine learning algorithms and optimization model to forecast greenhouse gas emissions with energy market data," *Sustainable Cities and Society*, vol. 82, Article ID 103886, 2022.

[65] J. Sun, T. Zhou, and D. Wang, "Relationships between urban form and air quality: a reconsideration based on evidence from China's five urban agglomerations during the COVID-19 pandemic," *Land Use Policy*, vol. 118, Article ID 106155, 2022.

[66] G. Chen, Y. Wang, S. Li et al., "Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: a satellite-based estimation using the random forests approach," *Environmental Pollution*, vol. 242, pp. 605–613, 2018.

[67] J. Ma, J. C. P. Cheng, F. Jiang, W. Chen, and J. Zhang, "Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques," *Land Use Policy*, vol. 94, Article ID 104537, 2020.

[68] Y. Liu, G. Cao, N. Zhao, K. Mulligan, and X. Ye, "Improve ground-level PM2. 5 concentration mapping using a random forests-based geostatistical approach," *Environmental Pollution*, vol. 235, pp. 272–282, 2018.

[69] Y. Tian, X. A. Yao, L. Mu, Q. Fan, and Y. Liu, "Integrating meteorological factors for better understanding of the urban form-air quality relationship," *Landscape Ecology*, vol. 35, no. 10, pp. 2357–2373, 2020.

[70] U. Grömping, "Variable importance assessment in regression: linear regression versus random forest," *The American Statistician*, vol. 63, no. 4, pp. 308–319, 2009.