The Influence Mechanism of Innovation Output in Resource-Based Cities from the Perspective of Eco-City: A Method Based on an Explainable Machine Learning Model

Liang Hong[1], Xiangfeng Li[2]

[1]Southeast University, Nanjing/China · 220240131@seu.edu.cn
[2]Southeast University, Nanjing/China

Abstract: Landscape environment plays a significant role in the innovation development of resource-based cities. From the perspective of eco-city, this study selects 23 regenerated resource-based cities in China as the research area and builds an interpretable machine learning method to uncover how service facilities, social economy, and the landscape environment relate to urban innovation output. Firstly, a reasonable factor framework was constructed. Secondly, the influencing factors were screened and models were compared. Thirdly, the contribution degree of factors was analysed using SHAP. Fourthly, the interaction effect of factors was analysed using PDP. This study reveals that science and education venues exert a significant positive effect on the innovation output of regenerated resource-based cities. High-quality landscape environments can effectively stimulate the innovation vitality of these venues. Explainable machine learning method can incorporate the landscape environment into the urban research framework, revealing the relationships between the landscape environment and various urban features. Through the interpretation of the results, targeted development suggestions can be provided for the city, which has strong practical guiding significance.

Keywords: Resource-based cities, eXtreme Gradient Boosting, Shapley Additive exPlanations, innovation output

1 Introduction

This is a descriptive paper of a complex process. Here, we present an explainable machine learning analysis method to reveal the mechanism between multiple influencing factors and urban innovation output. This method can incorporate multiple landscape environmental features into the urban research framework and analyse their synergy with other urban features.

Resource-based cities are those cities whose main development model is the exploitation and processing of natural resources. The single model driven by resource extraction has led to economic, social and environmental problems such as resource depletion and imbalanced industrial structure (FAN 2024). In 2013, China promulgated the Sustainable Development Plan of National Resource-based Cities, 2013-2023, classifying resource-based cities into four categories: growth, maturity, decline, and regeneration. Among them, the regenerated resource-based cities, as the forefront of innovative transformation among resource-based cities, have significant reference value for other cities in their pursuit of sustainable development (ZHENG et al. 2020). We select all the regenerated resource-based cities in China as the research subjects.

The current research on urban innovation development lacks explanations for the complex influencing mechanism. Most studies classify the driving factors of innovation development from one or two dimensions, but the understanding of how these factors affect each other and

Journal of Digital Landscape Architecture, 11-2026, pp. 108-118. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-770-0, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537770010.

their interrelationships is very limited. The landscape environment, as an important dimension of urban innovation development, is rarely taken into comprehensive consideration (ZHANG et al. 2023, ZHOU et al. 2025). Therefore, it is necessary to start from an overall perspective and construct a holistic framework to explore the influencing mechanism of multiple features to provide guidance for urban innovation development.

The concept of eco-city was initially proposed to address issues of industrial pollution and environmental deterioration. The introduction of the "ecological" concept from natural ecology emphasizes the need to conduct the overall design of cities from an ecological perspective to achieve harmonious development of society, economy and nature (CHEN et al. 2025). Currently, it's a key strategy for the development of resource-based cities.

To unravel the influence mechanism between urban innovation output and diverse factors, traditional quantitative models such as social network analysis, the super efficiency DEA model, and ordinary least squares analysis have been widely employed. However, it's difficult for these models to handle the nonlinear relationships that occur in real situations. Machine learning models offer an effective alternative for analyzing complex urban innovation. However, their application in research is limited by interpretability challenges. Explainable artificial intelligence, such as Shapley Additive exPlanations (SHAP), visualizes the influence mechanisms and can reveal the interrelationships between different factors.

Thus, this paper introduces an interpretable machine learning model. Among them, the Extreme Gradient Boosting (XGBoost)makes accurate predictions based on gradient boosting trees. SHAP model, based on the Shapley value in game theory, explains the importance of each factor. Partial Dependence Plots (PDP) describes the specific impact of individual or multiple factors on the model results. The interpretable machine learning model can deeply explain the influence mechanism of multiple factors, which is helpful for improving urban governance capabilities.

Here we provide a brief description of the interpretable machine learning method, emphasizing the roles of each step and the brief explanations of the diagrams. Through this work, we have reached three key conclusions.

1) The interpretable machine learning method, which integrates machine learning models, SHAP and PDP, can effectively analyse the nonlinear and interactive effects of multiple factors on urban innovation development, breaking through the limitations of traditional linear regression methods.

2) Explainable machine learning method can incorporate the landscape environment into the urban research framework, revealing the relationships between the landscape environment and various urban features. Through the interpretation of the results, targeted development suggestions can be provided for the city, which has strong practical guiding significance.

3) Science and education venues have a strong promoting effect on the innovation output of regenerated resource-based cities. High-quality landscape environments can effectively stimulate the innovation vitality of these venues.

2 A Method Based on an Explainable Machine Learning Model

2.1 Frame Construction

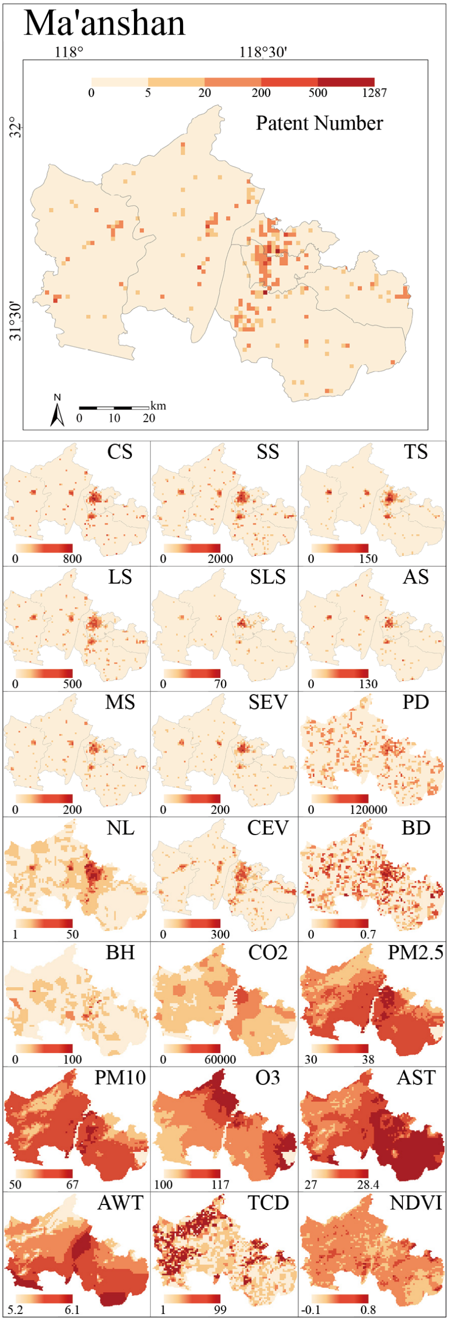


Fig. 1:
The spatial distribution of factors in Ma'anshan in 2023

The theory of eco-city emphasizes that cities are complex ecosystems integrating social, economic and natural elements. Its core lies in guiding the organization of urban space based on ecological principles. Landscape architecture focuses on the form and function of green spaces, emphasizing the improvement of urban environmental quality through systematic ecological design. In the process of innovative transformation of resource-based cities, they face dual challenges of ecological restoration and urban renewal, and need to coordinate and optimize the synergy of functional services, social economy and ecological environment. Therefore, this study, based on the theoretical connotations of eco-city and landscape architecture, has constructed a framework of three dimensions: service, society and environment. The service dimension focuses on public service facilities that provide diverse resources for innovation. The social dimension focuses on the development level of urban areas. The environmental dimension focuses on the ecological benefits of resource-based cities. Good air quality, comfortable climate, and high greenery quality can enhance the happiness and innovation ability of talents. The interactivity of these three dimensions can provide references for landscape architecture design.

2.2 Data Resource

Patent data is widely utilized by scholars for research on innovation in urban areas. Many researchers believe that although patent data cannot fully reflect all aspects of innovation, it remains one of the most direct, effective, and quantifiable indicators. The number of patents is generally regarded as an ideal quantitative indicator for measuring the innovation capacity of a region. Meanwhile, 21 factors related to urban innovation output were selected based on the literature and shown in Table 1. The service dimension includes catering, shopping, transportation, living, leisure, medical care, accommodation facilities. The social dimension includes population density, night light intensity, companies, building height and building density. The environmental dimension includes CO2, PM2.5, PM10, O3, summer average temperature, winter average temperature, tree coverage rate and normalized difference vegetation index (NDVI). The time range of the data in this study is 2013, 2018 and 2023. Spatial units are divided into grids of 1km by 1km. A 1km grid can precisely reflect the internal spatial characteristics and evolution patterns of a city at a micro scale and has been widely adopted in existing research. The 1km by 1km grid roughly corresponds to the blocks enclosed by main roads and is often regarded as a relatively complete urban spatial unit in terms of function. Research conclusions at this scale can be well applied to planning and construction practice. Due to page limitation, we have taken Ma'anshan City as an example to illustrate the spatial distribution of each factor in 2023 (Fig. 1).

Patent data are sourced from the China National Intellectual Property Administration. The night light intensity, NDVI and temperature data are obtained from National Earth System Science Data Center. The tree cover data are derived from the China Annual Tree Cover Dataset. POI data are chiefly derived from the Baidu Maps Open Platform. Air quality data are obtained from the China High Air Pollutants. The building density and population data are derived from the Global Human Settlement Layer, and the population data are corrected with census data. The carbon mission data are obtained from the Open-Data Inventory for Anthropogenic Carbon dioxide. The building height data are derived from the CMTBH-30 dataset.

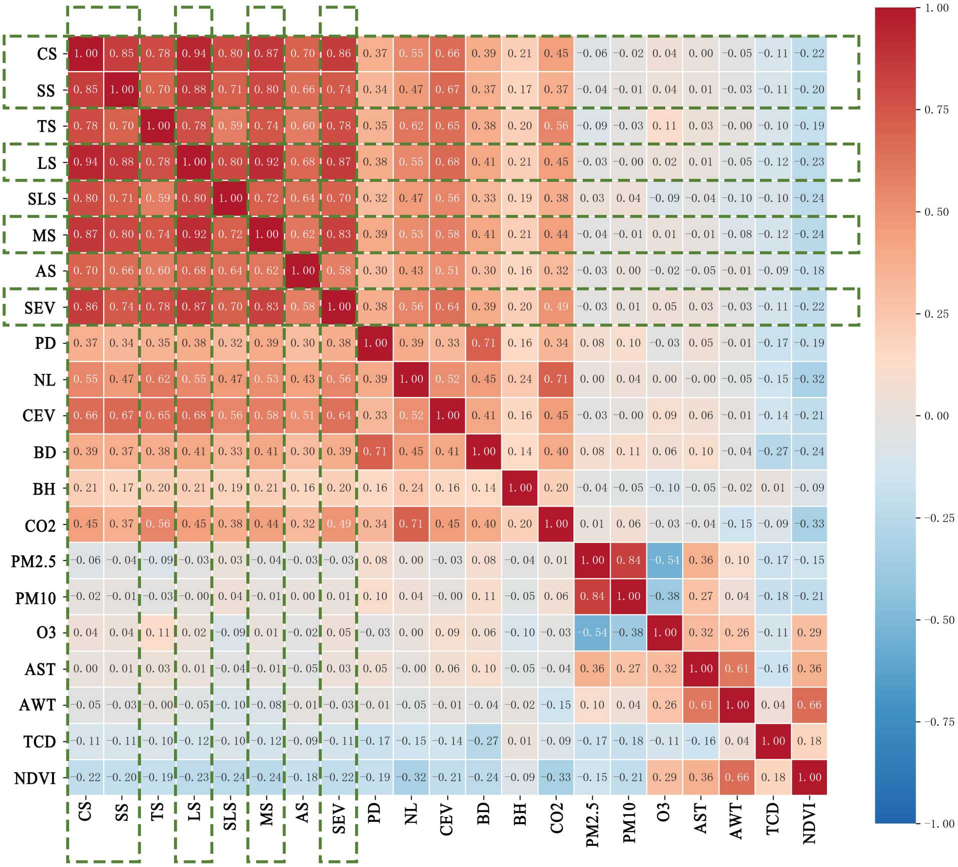
Table 1: Description of multidimensional influencing factors

Dimensions	Factors	Description
Service dimension	CS	The number of catering facilities (units/km^2)
	SS	The number of shopping facilities (units/km^2)
	TS	The number of transportation facilities (units/km^2)
	LS	The number of living facilities (units/km^2)
	SLS	The number of sports and leisure facilities (units/km^2)
	MS	The number of medical facilities (units/km^2)
	AS	The number of accommodation facilities (units/km^2)
Social dimension	PD	Number of resident population (persons/km^2)
	NL	Nighttime light intensity (nWcm-2 sr-1/km^2)
	CEV	Number of companies and enterprises (units/km^2)
	BD	Building density (%)
	BH	Average building height (m)
	SEV	The number of science and education venues (units/km^2)
Environmental dimension	CO2	Monthly fossil fuel carbon dioxide emissions (tons/km^2)
	PM2.5	Annual average PM2.5 concentration (μg/m^3)
	PM10	Annual average PM10 concentration (μg/m^3)
	O3	Annual average O3 concentration (ppb)
	AST	Average summer temperature (°C)
	AWT	Average winter temperature (°C)
	TCD	Tree cover degree (%)
	NDVI	Normalized Difference Vegetation Index

2.3 Factors Selection and Models Comparison

Highly correlated different factors often undermine the accuracy of the model and are usually regarded as redundant factors. The Pearson correlation coefficient is a statistical indicator for measuring the correlation between two variables and can be used for the preliminary screening of factors (Fig. 2). Notably, the correlations of CS, SS, LS, MS, and CEV all exceed 0.85. CS and LS belong to daily consumption facilities and have overlapping layouts with most service facilities. The VIF values of CS and LS are higher than 10, and in the initial analysis using machine learning, their SHAP values were found to be relatively low. In contrast, the istribution overlap of SS, MS, and SEV is relatively low, with VIF values less than 5 and higher SHAP values. Therefore, the two factors of CS and LS were excluded. As shown in Table 2, the VIF values of the remaining 19 factors are all less than 5, indicating that there is no multicollinearity in the model.

To demonstrate the applicability of XGBoost in this study, a total of 5 machine learning models and 2 traditional models were selected for comparative analysis. The machine learning models were Random Forest (RF), Adaptive Boosting (Adaboost), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Gradient Boosting Decision Tree (GBDT). The traditional models were Ordinary Least Squares (OLS) and Negative Binomial Regression (NB). For the machine learning model, 5-fold cross-validation is adopted to optimize the hyperparameters, reduce noise and enhance the generalization ability of the model. XGBoost outperforms other models with remarkable superiority and wide applicability. Compared to the linear assumption limitations of OLS, it can accurately capture the complex nonlinear relationships between variables, which is in line with the complex urban environment. Compared with SVR, it does not require cumbersome feature preprocessing and reduces the interference of outliers through regularization terms. Compared with random forests, it effectively suppresses overfitting through multiple mechanisms such as learning rate and regularization, and has better generalization performance. Compared with GBDT, with optimization techniques like parallel computing and sparse sensing, the training efficiency is significantly improved, and it is suitable for large scale data. Compared with NB, it has fewer restrictions on data types and distributions.



	CS	SS	TS	LS	SLS	MS	AS	SEV	PD	NL	CEV	BD	BH	CO2	PM2.5	PM10	O3	AST	AWT	TCD	NDVI
CS	1.00	0.85	0.78	0.94	0.80	0.87	0.70	0.86	0.37	0.55	0.66	0.39	0.21	0.45	-0.06	-0.02	0.04	0.00	-0.05	-0.11	-0.22
SS	0.85	1.00	0.70	0.88	0.71	0.80	0.66	0.74	0.34	0.47	0.67	0.37	0.17	0.37	-0.04	-0.01	0.04	0.01	-0.03	-0.11	-0.20
TS	0.78	0.70	1.00	0.78	0.59	0.74	0.60	0.78	0.35	0.62	0.65	0.38	0.20	0.56	-0.09	-0.03	0.11	0.03	-0.00	-0.10	-0.19
LS	0.94	0.88	0.78	1.00	0.80	0.92	0.68	0.87	0.38	0.55	0.68	0.41	0.21	0.45	-0.03	-0.00	0.02	0.01	-0.05	-0.12	-0.23
SLS	0.80	0.71	0.59	0.80	1.00	0.72	0.64	0.70	0.32	0.47	0.56	0.33	0.19	0.38	0.03	0.04	-0.09	-0.04	-0.10	-0.10	-0.24
MS	0.87	0.80	0.74	0.92	0.72	1.00	0.62	0.83	0.39	0.53	0.58	0.41	0.21	0.44	-0.04	-0.01	0.01	-0.01	-0.08	-0.12	-0.24
AS	0.70	0.66	0.60	0.68	0.64	0.62	1.00	0.58	0.30	0.43	0.51	0.30	0.16	0.32	-0.03	0.00	-0.02	-0.05	-0.01	-0.09	-0.18
SEV	0.86	0.74	0.78	0.87	0.70	0.83	0.58	1.00	0.38	0.56	0.64	0.39	0.20	0.49	-0.03	0.01	0.05	0.03	-0.03	-0.11	-0.22
PD	0.37	0.34	0.35	0.38	0.32	0.39	0.30	0.38	1.00	0.39	0.33	0.71	0.16	0.34	0.08	0.10	-0.03	0.05	-0.01	-0.17	-0.19
NL	0.55	0.47	0.62	0.55	0.47	0.53	0.43	0.56	0.39	1.00	0.52	0.45	0.24	0.71	0.00	0.04	0.00	-0.00	-0.05	-0.15	-0.32
CEV	0.66	0.67	0.65	0.68	0.56	0.58	0.51	0.64	0.33	0.52	1.00	0.41	0.16	0.45	-0.03	-0.00	0.09	0.06	-0.01	-0.14	-0.21
BD	0.39	0.37	0.38	0.41	0.33	0.41	0.30	0.39	0.71	0.45	0.41	1.00	0.14	0.40	0.08	0.11	0.06	0.10	-0.04	-0.27	-0.24
BH	0.21	0.17	0.20	0.21	0.19	0.21	0.16	0.20	0.16	0.24	0.16	0.14	1.00	0.20	-0.04	-0.05	-0.10	-0.05	-0.02	0.01	-0.09
CO2	0.45	0.37	0.56	0.45	0.38	0.44	0.32	0.49	0.34	0.71	0.45	0.40	0.20	1.00	0.01	0.06	-0.03	-0.04	-0.15	-0.09	-0.33
PM2.5	-0.06	-0.04	-0.09	-0.03	0.03	-0.04	-0.03	-0.03	0.08	0.00	-0.03	0.08	-0.04	0.01	1.00	0.84	-0.54	0.36	0.10	-0.17	-0.15
PM10	-0.02	-0.01	-0.03	-0.00	0.04	-0.01	0.00	0.01	0.10	0.04	-0.00	0.11	-0.05	0.06	0.84	1.00	-0.38	0.27	0.04	-0.18	-0.21
O3	0.04	0.04	0.11	0.02	-0.09	0.01	-0.02	0.05	-0.03	0.00	0.09	0.06	-0.10	-0.03	-0.54	-0.38	1.00	0.32	0.26	-0.11	0.29
AST	0.00	0.01	0.03	0.01	-0.04	-0.01	-0.05	0.03	0.05	-0.00	0.06	0.10	-0.05	-0.04	0.36	0.27	0.32	1.00	0.61	-0.16	0.36
AWT	-0.05	-0.03	-0.00	-0.05	-0.10	-0.08	-0.01	-0.03	-0.01	-0.05	-0.01	-0.04	-0.02	-0.15	0.10	0.04	0.26	0.61	1.00	0.04	0.66
TCD	-0.11	-0.11	-0.10	-0.12	-0.10	-0.12	-0.09	-0.11	-0.17	-0.15	-0.14	-0.27	0.01	-0.09	-0.17	-0.18	-0.11	-0.16	0.04	1.00	0.18
NDVI	-0.22	-0.20	-0.19	-0.23	-0.24	-0.24	-0.18	-0.22	-0.19	-0.32	-0.21	-0.24	-0.09	-0.33	-0.15	-0.21	0.29	0.36	0.66	0.18	1.00

Fig. 2: Pearson correlation heat map of influencing factors

As shown in Table 3, the R^2 values of the prediction set and the test set of XGBoost are 0.6923 and 0.6165 respectively. The 40% variance of the test set might be attributed to data noise or spatial heterogeneity. This indicates that the XGBoost model has higher scientific validity when analysing this dataset.

Table 2: VIF result of factors

Factor	**SS**	**TS**	**SLS**	**MS**	**AS**	**SEV**		
VIF	3.89	3.71	2.78	4.81	2.14	4.48		
Factor	**PD**	**NL**	**CEV**	**BD**	**BH**			
VIF	2.13	2.64	2.31	2.41	1.12			
Factor	**CO2**	**PM2.5**	**PM10**	**O3**	**AST**	**AWT**	**TCD**	**NDVI**
VIF	2.27	4.88	3.93	3.07	3.04	2.66	1.21	2.41

Table 3: Result of models (XGBoost key params: max_depth=5, learning_rate=0.01, n_estimators=1500, subsample=0.8, colsample_bytree=0.8, gamma=0.1, reg_alpha=1, reg_lambda=2, min_child_weight=2, random_state=2)

Model	RF	Adaboost	**XGBoost**	SVR	GBDT	OLS	NB
Prediction set R^2	0.5694	0.5726	**0.6923**	0.4439	0.5165	0.1555	0.2227
Test set R^2	0.3868	0.4948	**0.6165**	0.4116	0.3396		

3 Results and Discussion

3.1 The Relative Importance of Factors Explained by SHAP

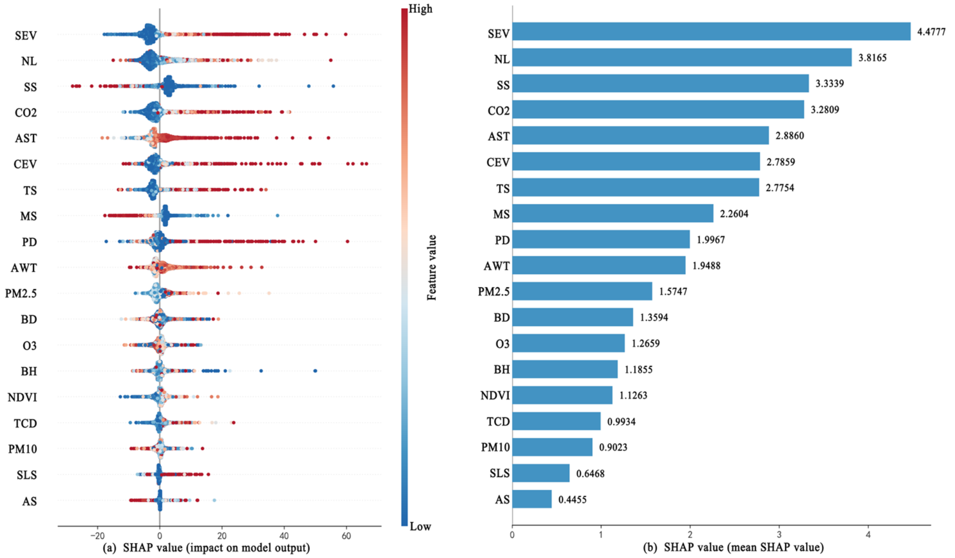


Fig. 3: Summary of relative importance of factors (SEV is Science and Education Venues)

To more accurately assess the influence degree of multiple factors, this study used the SHAP model to quantitatively rank the contribution degree of each factor (Fig. 3). In Figure 3a, the horizontal axis represents the SHAP value. Positive values indicate that the factor has a positive impact on the result, and the larger the value, the greater the contribution of the factor, and vice versa. Colours tending towards red indicate that the value of the factor is relatively high, while colours tending towards blue indicate that its value is relatively low.

As shown in Figure 3b, the relative importance of SEV is the highest (4.4777). The high value part is all greater than 0, and the low-value part is all less than 0. This indicates that the increase in urban science and education venues will lead to a significant increase in urban innovation output. Science and education venues provide a favourable cultural atmosphere and educational resources, attracting innovation spaces to settle in. Large meeting rooms offer venues for close communication among talents. Through cooperation among industry, academia, and research, the innovation output capacity is enhanced. For the environmental dimension, CO2 and AST are relatively more important. This might be because the materials for innovation activities in resource-based cities mainly consist of natural resources, which could lead to more CO2 pollution and high temperatures.

3.2 The Nonlinear Effect Based on PDP

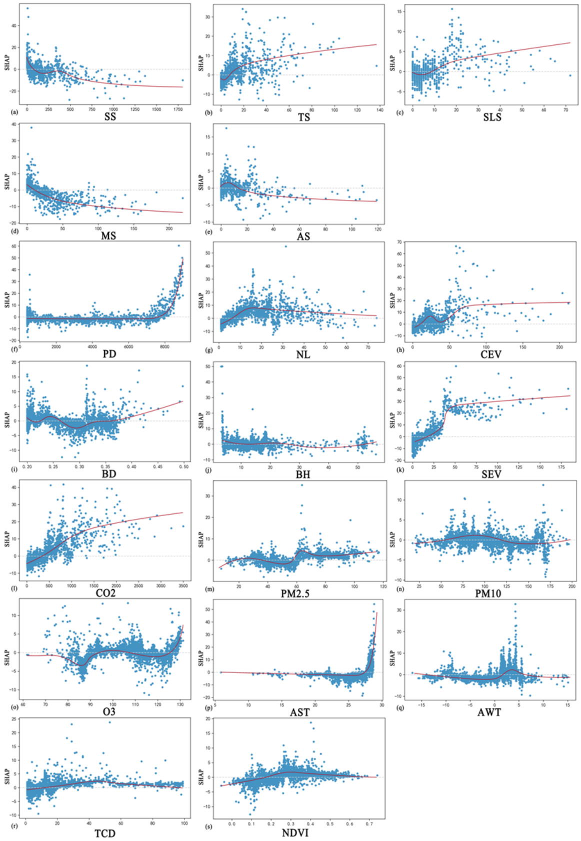


Fig. 4: The nonlinear effect of factors

This study further analysed the nonlinear relationship using the PDP model (Fig. 4). For SEV, the number of facilities ranging from 25 to 50 showed a significant increase in the promotion effect on innovation output, and then stabilized. This indicates that only when scientific and cultural facilities reach a certain degree of aggregation can they effectively promote regional innovation capabilities, and beyond the threshold, the improvement effect on innovation capabilities is weak. The influence of environmental dimension elements on innovation output can be classified into three categories. The first category is CO2. As its value increases, the overall impact shifts from inhibition to promotion, and within a certain range, the promoting effect rapidly intensifies. However, after reaching a certain threshold, the promoting effect gradually levels off. This can be explained by the fact that the research and development objects in resource-based cities with regeneration mainly focus on natural resources, which generates a large amount of CO2. However, the innovation capacity along this path is limited. The second category includes O3 and AST. High values overlap with innovation activity clusters, indicating that innovation activities are related to air pollution and high temperatures. The third category consists of PM2.5, PM10, AWT, TCD, and NDVI. As their values increase, the promoting effect on innovation output strengthens. However, after reaching a certain threshold, the promoting effect gradually weakens. This suggests that high-quality greenery can stimulate innovation vitality, but excessive greenery may occupy research and development space, leading to a decline in innovation vitality.

3.3 Interaction Effect of Multidimensional Influencing Factors

To gain a more comprehensive understanding of the interactive effects among the three dimensions, this study employed the SHAP feature interaction analysis and presents the interaction effects of 19 features with their most strongly interacting factors (Fig. 5). The horizontal axis represents the value of the feature, the vertical axis represents the interaction SHAP value, and the colour represents the value of the interacting feature. For example, the SHAP value of AST for patent output is positive, but the SHAP value of the interaction between AST and CEV is negative. This indicates that the summer heat and company clustering will reduce the promoting effect of summer heat on patent output, rather than suggesting that the summer heat and company clustering will inhibit patent output.

For the environmental dimension, the interaction between CO2 and PD is the strongest, and the interaction SHAP value is always negative, indicating that population concentration inhibits the promoting effect of CO2 on innovation output. When the AST value is high, as CEV increases, the negative value of the interaction SHAP value increases, suggesting that high summer temperatures inhibit the innovation capacity of areas with concentrated companies. The SHAP interaction value between AWT and O3 changes from negative to positive as AWT increases, indicating that O3 has significantly different effects on regions with different winter temperatures. When the SEV value is high, the interaction SHAP value between NDVI and SEV is high and increases with the increase in NDVI, suggesting that in areas with concentrated science and education venues, a better landscape environment can effectively promote innovation output. Innovation activities constitute cognitively demanding labour, and high-quality green spaces serve as restorative environments that alleviate mental fatigue. Concurrently, high-quality green public spaces provide venues for informal interactions among innovative talents, thereby effectively amplifying the spatial agglomeration effects of science and education venues. Therefore, eco-city needs to attach great importance to the construction of high-quality landscape environment around science and education venues.

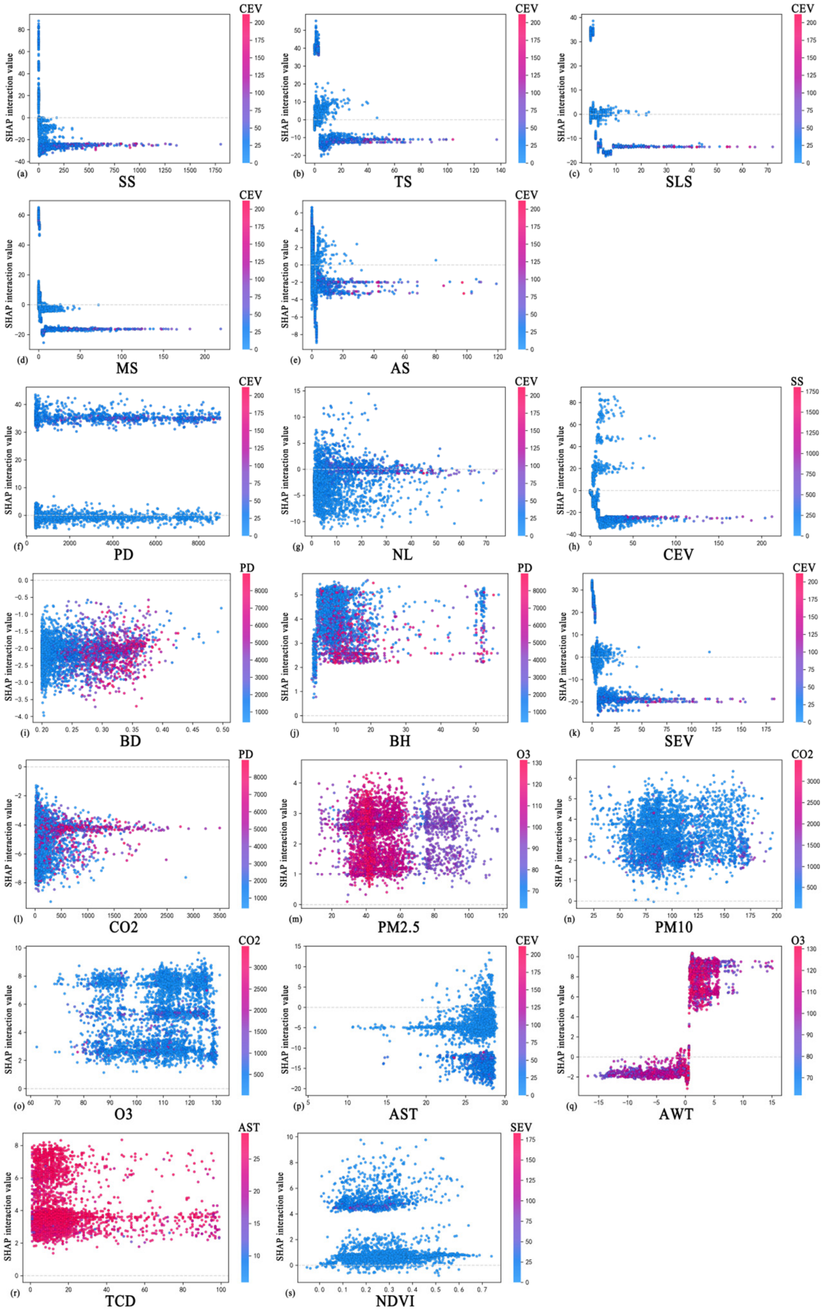


Fig. 5: The interaction effect of factors

4 Conclusion and Outlook

Landscape environment plays a significant role in the innovation development of resource-based cities. From the perspective of eco-city, this study selects 23 regenerated resource-based cities in China as the research area and builds an interpretable machine learning method to uncover how service facilities, social economy, and the landscape environment relate to urban innovation output. Firstly, a reasonable factor framework was constructed. Secondly, the influencing factors were screened and models were compared. Thirdly, the contribution degree of factors was analysed using SHAP. Fourthly, the interaction effect of factors was analysed using PDP.

Machine learning methods are being widely applied, but the relationship between the landscape environment and urban factors has not been given enough attention, and the interpretability is weak. We believe that interpretable machine learning methods should be used to analyse the role of the landscape environment in urban development. In summary, interpretable machine learning models can provide a unique perspective for landscape environment design and planning.

Although this study provides valuable method, it also has some limitations. Firstly, other features such as employment growth, R&D spending, and start-up formation could be incorporated into the innovation system. Secondly, innovation output has spatial autocorrelation, so local and global models could be constructed with using spatial weights to improve the accuracy of machine learning. Thirdly, due to paper constraints, we only briefly discussed the results. We will explore the reasons and propose suggestions in next paper.

References

CHEN, H., MHADHBI, M., TANG, R. & GOUTTE, S. (2025), Sustainable urban development policies and climate adaptation: evaluating real estate market stability in Tianjin Sino-Singapore Eco-City. Humanities and Social Sciences Communications, 12 (1). doi:10.1057/s41599-025-05627-9.

FAN, X. (2024), Research on the coupling of ecological environment and socio-economic development in resource-based cities: based on scenario simulation method. Sustainable Cities and Society, 114, 105810. doi:10.1016/j.scs.2024.105810.

ZHANG, J. X., CHENG, J. W., PHILBIN, S. P., BALLESTEROS-PEREZ, P., SKITMORE, M. & WANG, G. (2023), Influencing factors of urban innovation and development: a grounded theory analysis. Environment, Development and Sustainability, 25 (3), 2079-2104. doi:10.1007/s10668-022-02151-7.

ZHENG, Z., QIU, F. & ZHANG, X. (2020), Heterogeneity of correlation between the locational condition and industrial transformation of regenerative resource-based cities in China. Growth and Change, 51 (2), 771-791. doi:10.1111/grow.12365.

ZHOU, M., ZHOU, S., HEI, J., YANG, S., LIU, Q., YANG, T. & ZHIQIANG, W. (2025), Impact of innovation drivers in Chinese cities: machine learning analysis using XGBoost. Cities, 167, 106347. doi:10.1016/j.cities.2025.106347.