Research Paper

Modeling the effect of amenity determinants on urban innovation spaces using geographical random forest model

A Beijing case study

Tianheng SHU, Tsinghua University, China
Xia LIAO, Chongqing University, China
Shuo YANG, Tsinghua University, China
Taofang YU, Tsinghua University, China

Contact: Tianheng SHU, sthcqu@163.com

Abstract

Urban areas have different amenities that shape their innovation spaces. Previous research has examined urban amenities' effects at broader scales but rarely analyzes each factor's relative contribution and geographic variations. This study examines how different amenity factors affect urban innovation spaces in Beijing by using multiple global and local machine learning models at a 0.5 km² hexagonal grid scale. The analysis uses geographical random forests (GRF), a new nonlinear machine learning regression method, to examine how each amenity factor varies spatially and contributes to innovation spaces distribution. We then compare how GRF performs against geographically weighted regression (local model), random forest (global model), and ordinary least squares (global model). Our analysis shows that GRF works better than other models, and amenity factors vary in importance across different areas. The results show that education and research factor most strongly affect urban innovation spaces, followed by bus stations, healthcare amenity, and natural greenspace while recreation and entertainment amenity are relatively less significant. The influence of these factors also varies spatially; natural amenities and transportation accessibility exert a stronger influence in the outer areas, while commercial and cultural amenities are more pivotal in inner areas. Education and research facilities matter both in core areas and some outer regions. These patterns suggest that urban planning should develop specific strategies for different parts of the city to support innovation.

Keywords

Urban innovation spaces, Amenity determinants, Geographical random forest, Spatial analysis, Beijing

1. Introduction

Innovation is a key force in urban development, enhancing economic growth, advancing technology, and increasing a city's ability to compete globally and regionally (Johnson, 2008). In today's knowledge economy, innovation-focused cities not only modernize industries, but also attract talented individuals and

investment, transforming them into centers of creative activity and adaptability (Li and Du, 2022). This makes supporting innovation a major priority for urban policy and planning. Urban innovation spaces are the physical environments in which innovation occurs. These spaces include technology parks, incubators, and makerspaces that create opportunities for people to share ideas and resources and build networks that stimulate creativity. Through these spaces, people collaborate to create new products and services (Yigitcanlar, Adu-McVie and Erol, 2020). Beyond their physical presence, these spaces act as an ecosystem that connects people, businesses, and organizations, forming the foundation of the city's innovative capacity. Their growth depends on several factors, including market needs, policy support, and local amenities. Among these factors, urban amenities strongly influence the attractiveness and functionality of innovation spaces. Amenity determinants such as educational institutions, transportation, healthcare, and green spaces help create an environment where innovation can thrive (Arauzo-Carod, 2021). These amenities support innovation both directly and indirectly by enabling access to resources, reducing travel time, and improving the lives of those who work there.

In China, the study of urban innovation spaces is particularly important because cities such as Beijing, Shanghai, and Shenzhen are emerging as global innovation leaders (JLL, 2022). These cities illustrate China's move toward an innovation economy. To create effective policies for sustainable urban growth, we need to understand how innovation spaces work in these cities. It is important to investigate how different amenity determinants affect the development of innovation spaces in Chinese cities. This type of analysis can guide urban planning decisions by indicating the specific amenities that help innovation grow and concentrate in a particular area. By seeing how the effects of amenities vary geographically, planners can develop strategies that meet the needs of each inner urban space.

Previous research on urban innovation space has largely focused on the spatial clustering of innovation actors, especially the distribution of innovation districts and high-tech firms. Innovation districts and tech firms tend to be spatially concentrated, and these areas become centers for driving local economic development and attracting professionals (Bevilacqua, Parisi and Biancuzzo, 2019; Arauzo-Carod, 2021; Taecharungroj and Millington, 2022). Creative spaces also play an important role in urban innovation ecosystems by supporting cultural and artistic activities that enrich the overall innovation environment (He and Gebhardt, 2014). Analysis of these spaces often centers on their formation mechanisms and spatial distribution characteristics, which help us understand the processes and factors that influence the development of urban innovation spaces (Yigitcanlar, Adu-McVie and Erol, 2020; Yuan *et al.*, 2021; Tan, Qian and Chen, 2023).

While much attention has been given to the structural aspects of innovation spaces, relatively few studies look at the effects of amenity determinants on these spaces. Studies about urban amenity often use qualitative methods like case studies to examine how amenity factors attract and support innovation spaces (Esmaeilpoorarabi *et al.*, 2018; Adu-McVie *et al.*, 2021). Quantitative research typically uses traditional linear econometric models - including correlation analysis (Taecharungroj and Millington, 2022), count data models (Arauzo-Carod, 2021), and geographically weighted regression (GWR) (Zhou and Yang, 2024) - to study the connections between amenity determinants and innovation spaces. However, linear models may oversimplify the complex non-linear relationships and spatial variations between these elements, making it harder to understand deeply how different amenity determinants shape innovation spaces.

Recent advancements in modeling techniques, such as the geographical random forest (GRF) model, represent new approaches that addresses these modeling challenges. Developed by Georganos et al. (2021), GRF combines the benefits of machine learning with GWR, effectively capturing local variations and

non-linear effects within spatial data. This model is increasingly used in urban studies to address spatial heterogeneity and shows promise in revealing complex patterns and local interactions that may be missed by traditional linear models (Grekousis *et al.*, 2022; Lotfata and Georganos, 2023; Lotfata, Grekousis and Wang, 2023; Wang *et al.*, 2024).

To conclude, while extant studies have made progress in identifying the general factors that contribute to urban innovation spaces, research that systematically examines the interaction between various amenity determinants on the spatial distribution of these spaces remains limited. Given the potentially complex and nonlinear relationships, traditional linear models may not fully capture these dynamics. This study addresses these gaps by employing GRF to provide a more detailed analysis of how amenity determinants influence the location and density of innovation spaces within a city, offering new insights into the spatial heterogeneity of these effects.

Given these considerations, this study aims to answer the following questions: How do various amenity determinants contribute to the distribution of innovation spaces within a city? and is there local heterogeneity in how these determinants impact innovation spaces?

2. Method and data

Figure 1 shows our research framework, which has three main parts. First, we analyze data and variables. Second, we build and compare different models. Third, we examine the key results. We use the GRF model to study how amenity determinants affect where urban innovation spaces are located. This model helps us understand these relationships across different areas and how they interact with each other.

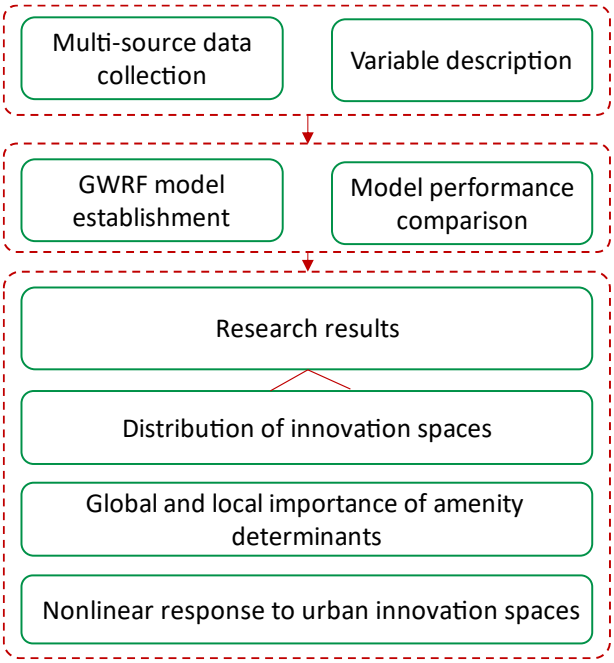

Figure 1. Workflow of this study. Source: Drawn by the authors

2.1. Variable description and data sources

This study aims to explore the impact of various amenity determinants on urban innovation spaces. To achieve this, we first need to measure innovation spaces within the city. Innovation space represents the physical area where innovation-dominant entities are clustered in urban microspatial units. In this study, we focus on key innovative enterprises, such as high-tech enterprises and innovative small and medium-sized enterprises (SMEs) that have a tendency to form spatial clusters. The choice of spatial location for these firms is influenced by amenity factors; for instance, enterprises often prefer locations with convenient transportation.

To quantify innovation spaces, we measure the density of these innovative enterprises within each study unit to understand both the clustering and spread of innovation activities. This variable is denoted as $kd_innovation$, and is calculated by kernel density estimation of the innovative enterprises within each spatial unit. Data on these innovative enterprises are provided by Qichacha, a Chinese enterprise information query platform. We further extracted enterprise addresses and used the Amap API for geocoding (Amap, 2024), enabling spatial positioning within the city.

Following existing literature on urban amenities (Ullman, 1954; Glaeser, 1999; Florida, 2014; Taecharungroj and Millington, 2022), we selected five categories of amenity determinants: natural factors, transportation accessibility, cultural factor, commercial factors, and education & research factor. Each category contains one or more indicators as shown in Table 1. Most of the amenity determinants data come from Amap's Points of Interest (POI) dataset. Additionally, greenspace data is sourced from high-resolution (1m) green space data published by Shi *et al.* (2023).

Table 1 Variable description. Source: Compiled by the authors

Amenity Type	Variable	Description	Unit
Natural factors	$greenspace$	Green space area	km^2
	$distance_to_park$	Distance to nearest park	m
Transportation accessibility	$distance_to_subway$	Distance to subway station	m
	$distance_to_highspeed$	Distance to highspeed rail station	points/km^2
	$kd_bus_station$	Kernel density for bus station	points/km^2
Cultural factor	$kd_culture$	Kernel density for cultural facility POI	points/km^2
Commercial factors	$kd_shopping$	Kernel density for shopping POI	points/km^2
	kd_daily_life	Kernel density for daily life POI	points/km^2
	$kd_recreat_enter$	Kernel density for recreation and entertainment POI	points/km^2
	$kd_healthcare$	Kernel density for healthcare POI	points/km^2
	kd_food_drink	Kernel density for food and drink POI	points/km^2
	$kd_finance_insurance$	Kernel density for finance and insurance POI	points/km^2
Education and research factor	$kd_edu_research$	Kernel density for education and research POI	points/km^2

As shown in Figure 2, when we map these factors, we can see clear differences across different areas. Each type of amenity has its own unique spatial pattern. Some areas have better access to resources, while others have limited availability.

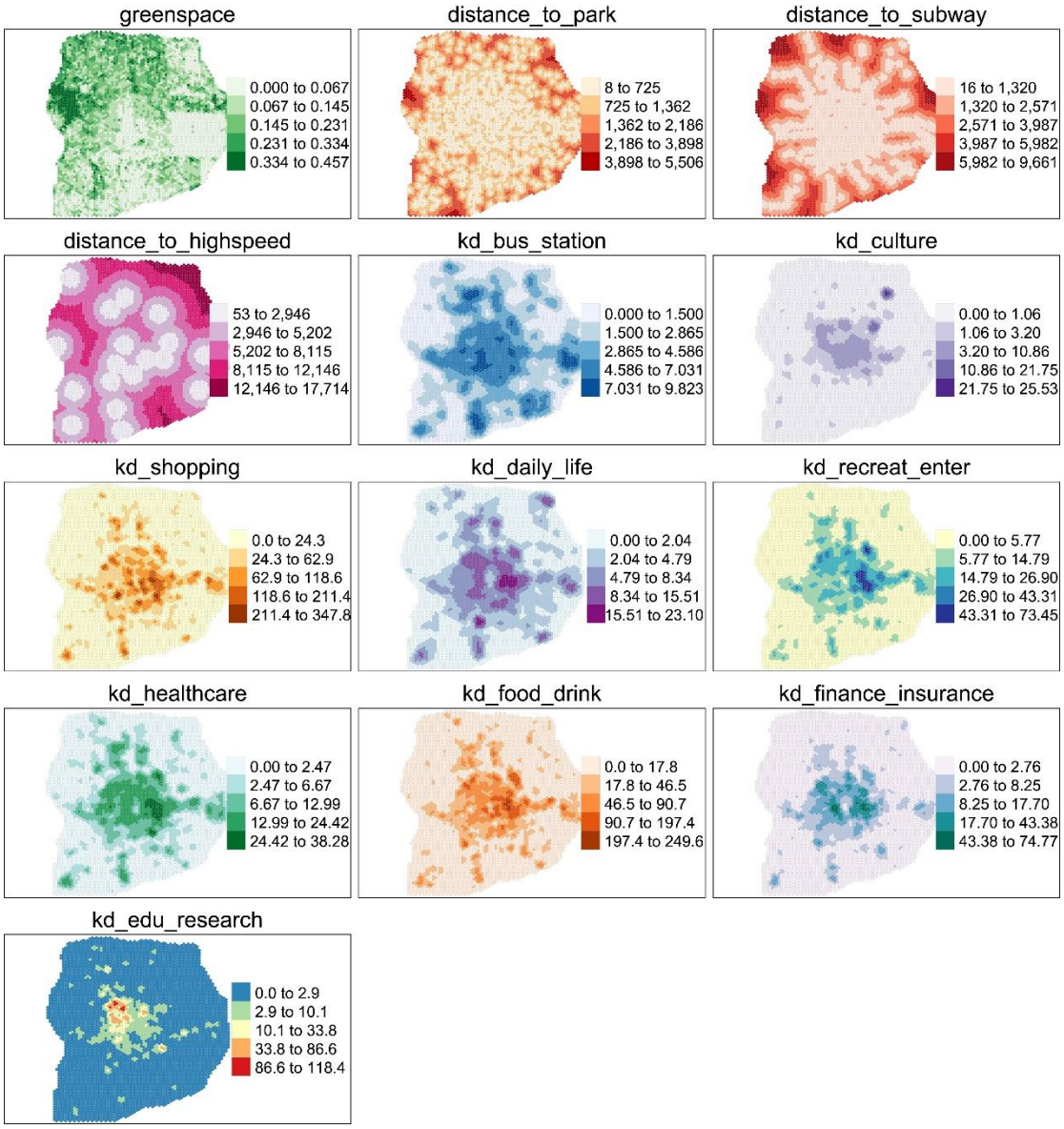

Figure 2. Spatial variation of amenity determinants. Source: Drawn by the authors

2.2. Development of GRF Model

GRF model is an extension of the traditional Random Forest (RF) model, designed to incorporate spatial heterogeneity into predictions, making it particularly useful in urban studies where spatial effects are significant (Georganos *et al.*, 2021). Unlike traditional RF, which assumes a single global model for all data points, GRF applies multiple local models to account for spatial variability, similar to the approach used in GWR.

The core idea behind GRF is to leverage local variations by calibrating a separate RF model for each spatial location, rather than relying on a single model across the entire study area (Georganos *et al.*, 2021; Georganos and Kalogirou, 2022). This approach allows GRF to capture the localized relationships between independent variables and the dependent variable, improving model accuracy in scenarios where spatial heterogeneity plays a crucial role. In GRF, each local model is calibrated using a spatial weights matrix and a subset of observations that are spatially proximate to the location being analyzed. For each spatial unit i, a separate RF model is trained using only nearby data points within a specified spatial window, defined by a distance or neighborhood threshold. This approach ensures that predictions reflect local conditions, allowing the model to adapt to variations in the spatial distribution of independent variables.

To better understand GRF, we first outline the standard RF model. In a traditional RF model, the dependent variable y_i at location i is predicted based on a nonlinear function of the independent variables ax_i, with an error term ε:

$$y_i = ax_i + \varepsilon \quad (1)$$

The GRF model extends the RF framework by introducing local calibrations, allowing predictions to account for spatial heterogeneity. The GRF prediction equation at each location i can be represented as follows:

$$y_i = a(u_i, v_i)x_i + \varepsilon \quad (2)$$

where $a(u_i, v_i)$ represents the RF model calibrated specifically for location i, with (u_i, v_i) denoting the spatial coordinates of the centroid of the spatial unit i. This location-specific calibration uses only a subset of the original dataset, defined by a spatial window around each point. To determine the subset of observations used in each local model, GRF applies a spatial weighting function, typically based on a kernel density approach. A common choice is the adaptive kernel, where the size of the kernel adjusts to include a fixed number of nearest neighbors around each observation. This method is similar to the approach used in GWR and allows for flexibility in areas with varying data densities.

2.3. Study area and units

We chose Beijing as the research area for several reasons. As China's political, cultural and technological center, Beijing has a high concentration of innovative enterprises, making it an ideal setting for analyzing the spatial distribution of innovation spaces. Furthermore, Beijing's status as a major innovation center ensures a rich dataset of various amenity determinants and innovation activity, providing an opportunity to explore the complex relationships between them. We limited our research to the area inside Beijing's Sixth Ring Road, where most economic activities happen (Figure 3a). The city's main business operations take place in this central area. In contrast, areas outside the Sixth Ring Road have few innovative enterprises, and in some places, none at all. If we included these outer areas in our study, it would make our model predictions less accurate because innovation spaces there are scattered and hard to define.

For a fine-grained spatial analysis, we divided the area into hexagonal cells. Each cell is 0.5 km² (Figure 3b), about the same size as a typical block in Beijing's core districts (People's Government of Beijing Municipality, 2018). Hexagons provide consistent measurement units across the study area, avoiding size and shape variability issues associated with administrative boundaries. This structure reduces edge effects, enhances spatial comparability, and minimizes boundary biases in analyzing the distribution of innovation spaces and amenities (Zhang, Xiong and Wei, 2022; Liu, Meng and Shi, 2023).

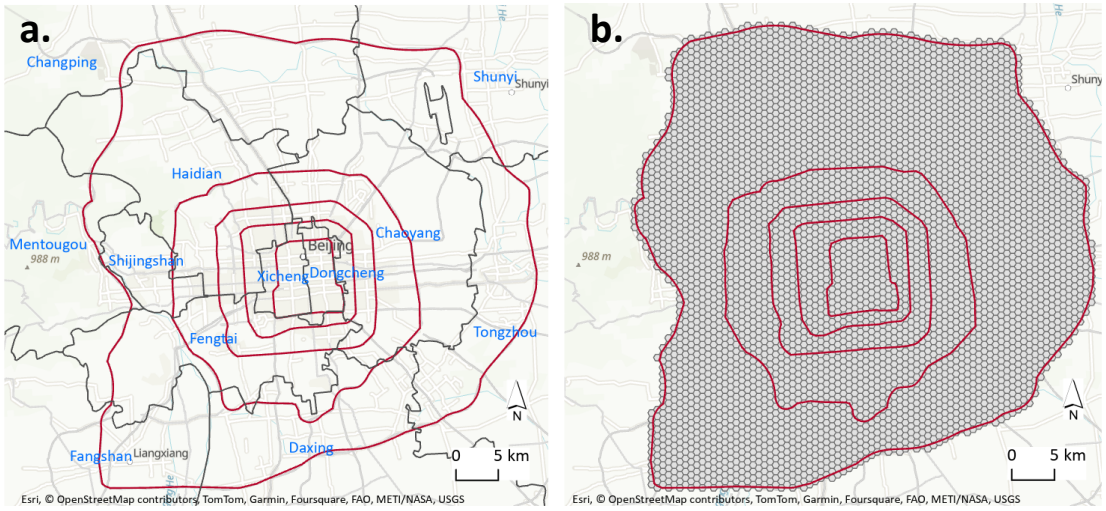

Figure 3. Study area (a) and spatial units (b). Source: Drawn by the authors

2.4. Model performance metrics and model preparation

To assess the performance of our model, we use three common evaluation metrics: R-square (model fitness), mean square error (MSE), and mean absolute error (MAE). These metrics enable to verify how well the model captures the relationships within the data and its predictive accuracy. The formulas for each metric are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

where where y_i represents the observed values, $\hat{y}_i$ represents the predicted values, and $\bar{y}$ is the mean of the observed values.

Before implementing the GRF model, several preparatory steps were taken to ensure the reliability and robustness of the results. First, a co-linearity check was performed to investigate potential multicollinearity among the amenity determinants. By calculating the variance inflation factor (VIF) for each variable, we ensured that all VIF values were less than 5, indicating that collinearity would not harmfully impact the model's performance or interpretation.

With the amenity determinants confirmed for inclusion, hyperparameter tuning was conducted to optimize the model's performance. The process involved a random grid search using the "caret" package in R. Various configurations of the parameters of the random forest were tested. A 10-fold cross-validation approach was applied and parameter combinations were iterated to identify optimal settings. The results

showed that the best performance was achieved with 500 trees (ntree=500) and 8 randomly selected variables in each split (mtry=8).

Finally, bandwidth selection was necessary to define the local neighborhood size of the GRF model. The Golden Section Search method was applied to minimize the corrected Akaike information criterion (AICc) and to identify the optimal bandwidth. This method determines a bandwidth of 68, which allows the model to balance local sensitivity with spatial smoothness and effectively capture spatial heterogeneity without over-fitting.

3. Results and analysis

3.1. Spatial pattern of innovation spaces in Beijing

We first marked all innovative enterprises on a map. Then we used different methods to analyze how these enterprises spread across Beijing. Our analysis methods included kernel density estimation, grid aggregation, and LISA clustering analysis (Figure 4). The results showed that these enterprises tend to form clusters in certain areas, showing a polycentric distribution across the city.

The major clusters identified include Zhongguancun, Shangdi, CBD (Central Business District), Wangjing, and Yizhuang. These clusters demonstrate that Beijing's innovation spaces are not randomly dispersed, but strategically concentrated in specific areas known for high economic activity, robust infrastructure, and a wealth of amenities that support innovation. This polycentric distribution reflects a deliberate spatial organization in which multiple centers of innovation emerge throughout the city, perhaps influenced by a variety of amenity determinants and urban policies tailored to each region.

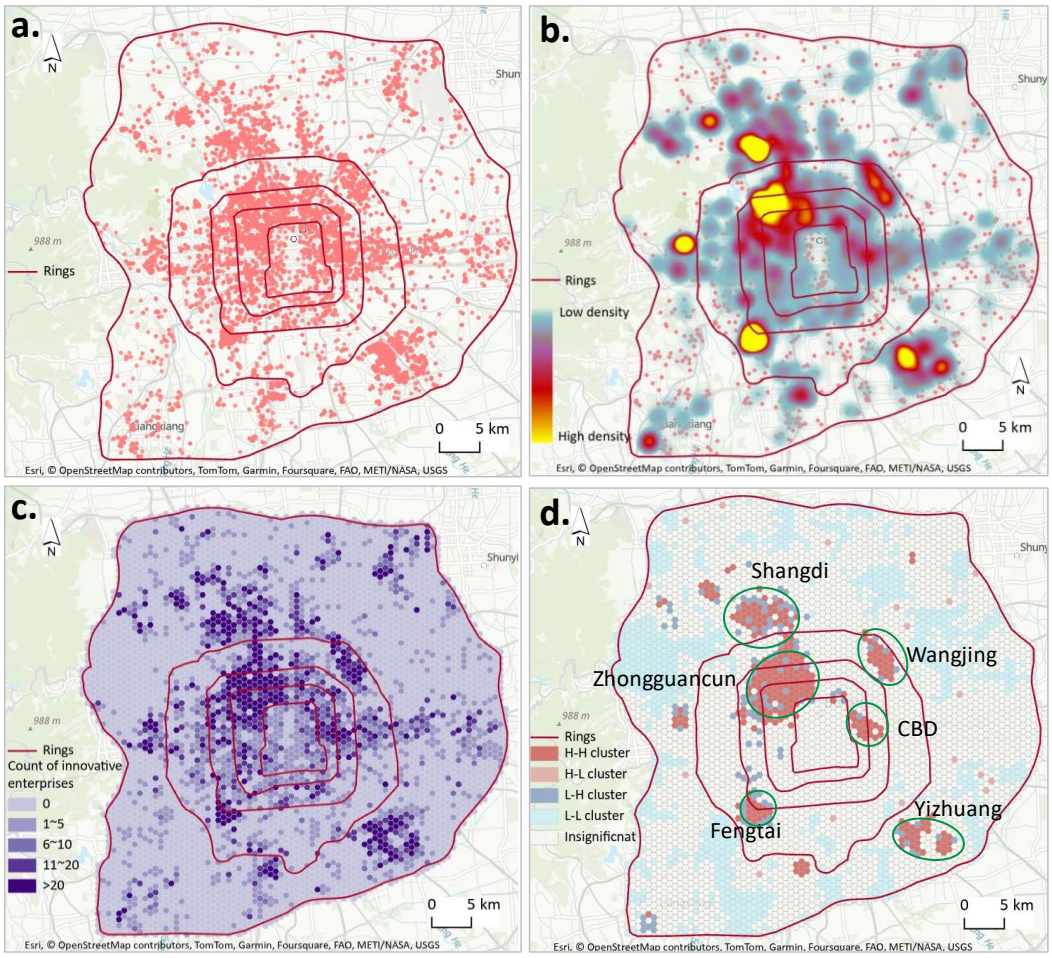

Figure 4. Spatial distribution of urban innovation spaces in Beijing. a, Location of innovative enterprises. b, Kernel density distribution of innovative enterprises. c, Count of innovative enterprises. d, Anselin Local Moran's I clusters (see method in Anselin, 1995). Source: Drawn by the authors

3.2. Model performance comparison

We tested four different models in this study: the GRF model, the global ordinary least squares (OLS) model, the traditional RF model, and the local GWR model. To see which model worked best, we looked at their R^2, MSE, and MAE values, as we mentioned earlier. The GRF model got better results than all other models. It had the highest R^2 at 0.909, and the lowest MSE (76.40) and MAE (3.19) values (see Table 2). The GRF model's high accuracy comes from its ability to capture both nonlinear relationships and spatial heterogeneity in the data. These two features make the model effective for studying how innovation spaces spread across different parts of a city.

Table 2. Metric performance of different models. Source: Compiled by the authors

Model	R^2	MSE	MAE
OLS	0.290	595.25	11.02
RF	0.682[a]	266.8	6.92
GWR	0.853	122.94	5.52
GRF	0.909[a]	76.40	3.19

[a] Metrics are computed from Out-of-Bag samples

3.3. Importance of amenity determinants' effect on innovation spaces

Using the "ranger" package in R, we ran a global RF model to determine the relative importance of various amenity determinants in predicting the distribution of $kd_innovation$. The importance of each factor was evaluated based on the increase in Mean Squared Error (IncMSE) when each amenity determinant was randomly permuted (Breiman, 2001), meaning that a higher IncMSE indicates a more important factor for the model's predictive power.

As shown in Figure 5, education and research amenity is the most important factor influencing the spatial distribution of Beijing's urban innovation space. This suggests that proximity to universities, research centers, and other educational facilities strongly attracts innovative enterprises, possibly because of the availability of skilled human resources and research partnerships in these areas (Oh *et al.*, 2016). Following education and research amenity, bus station density ranks as the second most important factor, highlighting the role of transportation accessibility in supporting innovation spaces. Healthcare amenity and natural greenspaces are also important elements, which suggests innovative enterprises prefer areas with good health facilities and natural environments. In contrast, determinants related to recreation and entertainment and proximity to parks show relatively lower importance. While these determinants can

make urban life better overall, they don't seem to strongly influence where innovative enterprises choose to locate in Beijing.

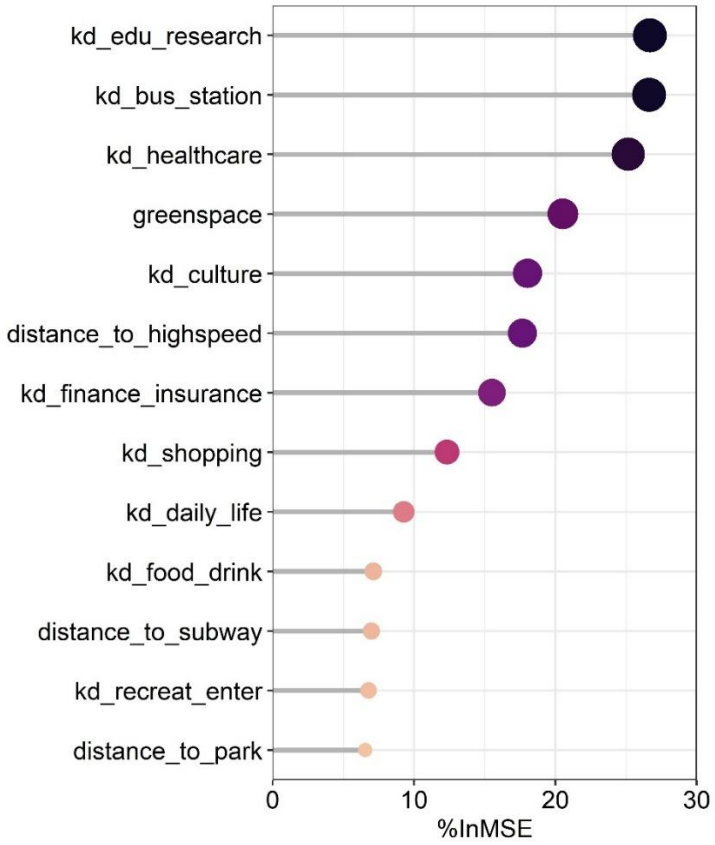

Figure 5. Global importance of amenity determinants. Source: Drawn by the authors

Furthermore, using the "SpatialML" package in R, we applied a local GRF model to analyze the spatial variability in the importance of different amenity determinants across Beijing's urban zones. The results reveal distinct patterns: in the outer areas, natural factors like greenspace and transportation accessibility (proximity to subway and high-speed rail) are highly influential. This makes sense because companies in these areas need both a pleasant working environment and quick ways to reach the city center. In contrast, central areas depend more on cultural, commercial, and transit-related amenities. Cultural facilities and daily life services add vibrancy and convenience, appealing to talent in dense urban environments, while bus station density enhances connectivity within these areas. Healthcare, shopping, and food amenities are also important, as they help workers' daily lives and make business operations smoother for innovative enterprises.

A unique factor, education and research factor, holds substantial influence in both core and specific peripheral regions. In the central city, they provide direct access to talent and research resources, while in outer zones, they help create new clusters of innovative companies, showing how important educational and research facilities are for innovation across the whole city.

In summary, the spatial patterns reveal that outer zones emphasize natural factors and transportation accessibility, while central areas rely more on cultural, commercial, and transit-related dterminants. Education and research factor hold unique significance across both core and certain peripheral regions, acting as foundational elements for innovation. Based on these findings, city planners should develop

different strategies for different areas when placing public facilities. This way, they can create better conditions for innovation spaces across the entire city.

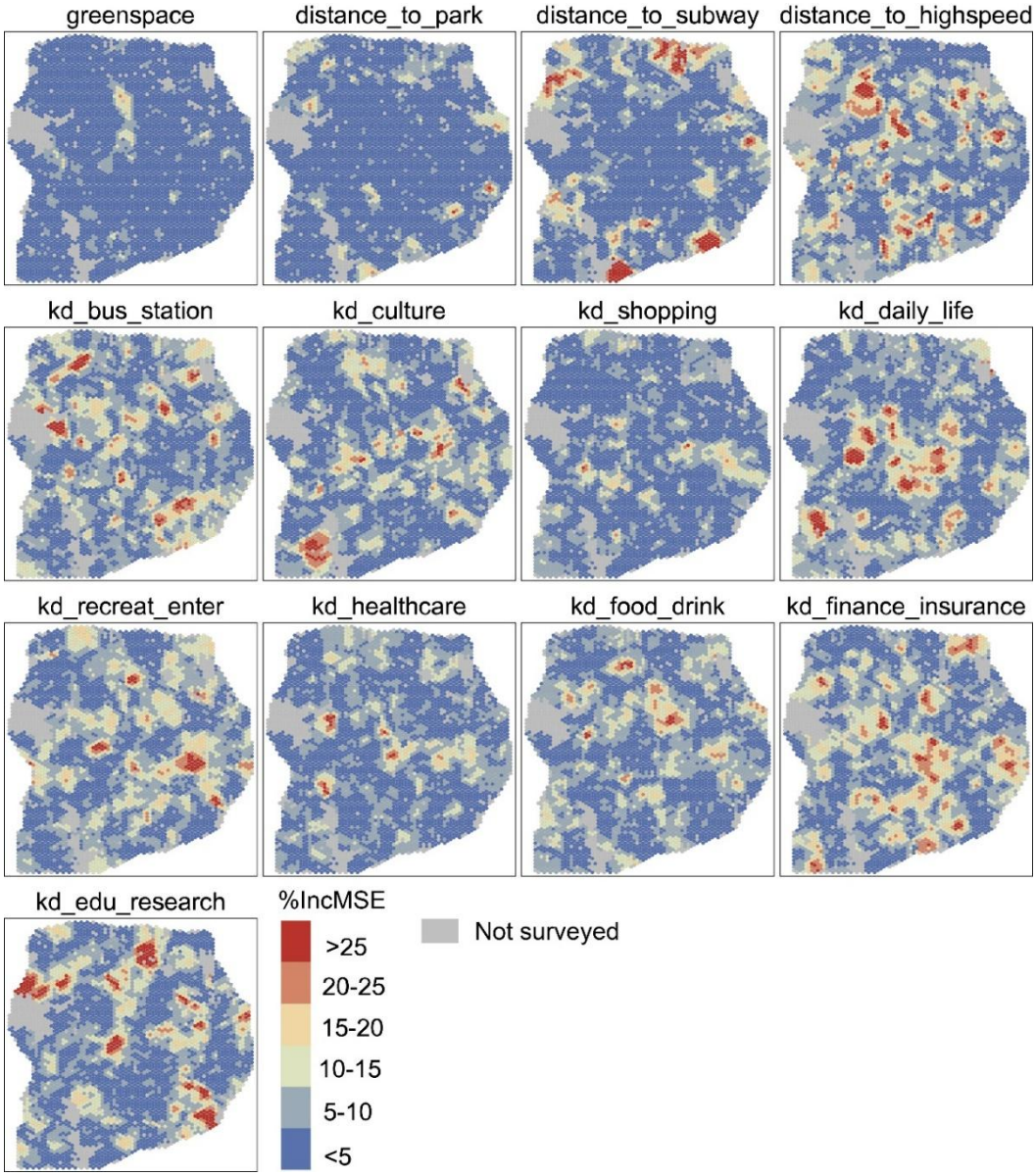

Figure 6. Local importance of amenity determinants. Source: Drawn by the authors

3.4. Nonlinear response of amenity determinants to innovation spaces

To explore the nonlinear effects of amenity determinants on innovation spaces, we analyzed the response curves of selected amenity determinants on $kd_innovation$ using the "spatialRF" package in R. These response curves illustrate each predictor's effect on $kd_innovation$ while holding other predictors constant at their median values. The RF model was repeated 30 times, with the red lines representing the median one (Figure 7).

The response curves show that many predictors exhibit nonlinear threshold effects and that after a certain point, their impact on kdd _innovation stabilizes. Below are some key examples of each facility type. Here are some key examples for each amenity type:

Green space ($greenspace$): when the green space area reaches about 0.3 km^2, the impact on $kd_innovation$ increases rapidly and then levels off. This indicates that green space enhances the surrounding innovation space, but the impact of additional green space above this threshold is limited.

Distance to highspeed rail station ($distance_to_highspeed$): A U-shaped response shows that $kd_innovation$ increases as distance to high-speed rail stations decreases, reaching an optimal point around 5,000 meters. However, locations too close to stations might have less innovation, probably because of noise and traffic near these stations.

Cultural amenity ($kd_culture$): kd_innovation increases with cultural facilities up to 10 nearby locations, then stops growing. This means a diversified cultural environment encourages innovation, but more facilities add little value.

Healthcare amenity ($kd_healthcare$): $kd_innovation$ grows with the increase of healthcare facilities in the neighborhood, peaking around 12. This suggests that moderate access to healthcare supports innovation spaces without oversaturation.

Education and Research ($kd_edu_research$): $kd_innovation$ benefits from educational and research institutions up to around 15, then levels off, suggesting a moderate clustering of these resources is optimal.

These nonlinear response patterns indicate a threshold effect, indicating that while conveniences such as education, culture, and health care intensify $kd_innovation$, their impact stabilizes after a certain level. For this reason, balanced comfort measures are more effective than oversaturation.

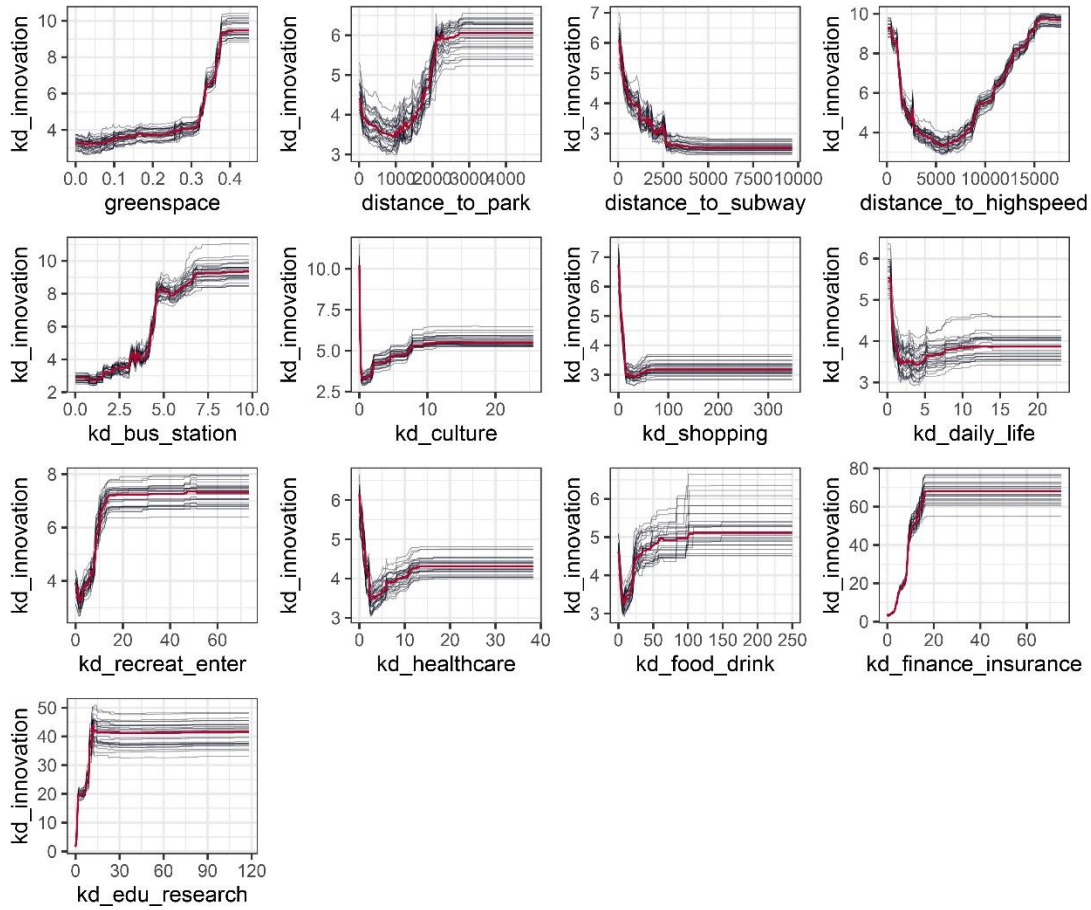

Figure 7. Response curves of amenity determinants. Source: Drawn by the authors

4. Conclusion and implications

This study employed the GRF model to assess the complex, spatially heterogeneous effects of various amenity determinants on the distribution of urban innovation spaces in Beijing. By leveraging the GRF model, we were able to account for both the nonlinear relationships and spatial heterogeneity inherent in the influence of amenity determinants on innovation spaces. Our findings reveal that education and research factor are the most influential determinant, demonstrating their importance in promoting innovation by offering access to skilled talent and knowledge resources. Additionally, the impact of different amenity determinants differs between core and peripheral areas, underlining the importance the importance of spatial context in understanding amenity effects. The nonlinear response patterns observed—particularly threshold effects and saturation points—suggest that urban amenities can support innovation spaces up to an optimal level, beyond which additional amenities contribute less significantly.

Theoretically, this research advances our understanding of the spatial dynamics of urban innovation spaces by integrating the effects of multiple amenity determinants through a localized, nonlinear analytical approach. Our findings demonstrate that the spatial distribution of innovation spaces is determined by both the type and concentration of surrounding amenities, adding new insights to urban studies literature.

For urban planning practice, our research has direct implications. Urban planners must consider both spatial heterogeneity and nonlinear effects when developing amenity provision strategies. Recognizing the

varying requirements of innovation spaces across different city zones, planners can tailor the distribution and intensity of key amenities—such as education facilities in both core and strategic outer areas, and healthcare and cultural services in the central zones—to optimize their effectiveness in attracting innovative enterprises. This targeted approach can enhance the urban environment's appeal to innovative entities, supporting a more balanced and innovation-friendly city structure.

This research has several limitations that point to directions for future study. First, we only looked at innovation spaces at one point in time. A long-term analysis would show how these spaces change over time, helping us understand their development patterns. Additionally, future studies could explore the interactive effects of different amenity determinants on innovation spaces to refine amenity provision strategies further. Finally, our study treated all innovation spaces the same way. Investigating how amenity determinants affect different types of innovation spaces - like those in service industries versus manufacturing - would help planners better meet the specific needs of each sector.

5. Acknowledgements

This work was supported by National Key Research and Development Program of China with grant number 2022YFC3800101. The authors acknowledge the use of ChatGPT-4 (OpenAI, 2023) for language improvement in this paper.

6. References

Adu-McVie, R. *et al.* (2021) 'Classifying innovation districts: Delphi validation of a multidimensional framework', *Land Use Policy*, 111, p. 105779. Available at: https://doi.org/10.1016/j.landusepol.2021.105779.

Amap (2024) *Geododing-Amap API*. Available at: https://lbs.amap.com/api/webservice/guide/api/georegeo (Accessed: 18 September 2024).

Anselin, L. (1995) 'Local Indicators of Spatial Association—LISA', *Geographical Analysis*, 27(2), pp. 93–115. Available at: https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Arauzo-Carod, J.-M. (2021) 'Location determinants of high-tech firms: an intra-urban approach', *Industry and Innovation*, 28(10), pp. 1225–1248. Available at: https://doi.org/10.1080/13662716.2021.1929868.

Bevilacqua, C., Parisi, L. and Biancuzzo, L. (2019) 'Multi-stage Strategic Approach in Spatial Innovation: How Innovation District Matter?', in F. Calabrò, L. Della Spina, and C. Bevilacqua (eds) *New Metropolitan Perspectives*. Cham: Springer International Publishing, pp. 85–94. Available at: https://doi.org/10.1007/978-3-319-92099-3_11.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/A:1010933404324.

Esmaeilpoorarabi, N. *et al.* (2018) 'Does place quality matter for innovation districts? Determining the essential place characteristics from Brisbane's knowledge precincts', *Land Use Policy*, 79, pp. 734–747. Available at: https://doi.org/10.1016/j.landusepol.2018.09.016.

Florida, R. (2014) 'The Creative Class and Economic Development', *Economic Development Quarterly*, 28(3), pp. 196–205. Available at: https://doi.org/10.1177/0891242414541693.

Georganos, S. *et al.* (2021) 'Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling', *Geocarto International*, 36(2), pp. 121–136. Available at: https://doi.org/10.1080/10106049.2019.1595177.

Georganos, S. and Kalogirou, S. (2022) 'A Forest of Forests: A Spatially Weighted and Computationally Efficient Formulation of Geographical Random Forests', *ISPRS International Journal of Geo-Information*, 11(9), p. 471. Available at: https://doi.org/10.3390/ijgi11090471.

Glaeser, E.L. (1999) 'Learning in Cities', *Journal of Urban Economics*, 46(2), pp. 254–277. Available at: https://doi.org/10.1006/juec.1998.2121.

Grekousis, G. *et al.* (2022) 'Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach', *Health & Place*, 74, p. 102744. Available at: https://doi.org/10.1016/j.healthplace.2022.102744.

He, J.-L. and Gebhardt, H. (2014) 'Space of Creative Industries: A Case Study of Spatial Characteristics of Creative Clusters in Shanghai', *European Planning Studies*, 22(11), pp. 2351–2368. Available at: https://doi.org/10.1080/09654313.2013.837430.

JLL (2022) *Beijing, Shanghai, and Shenzhen listed in Global Top 15 Innovative Markets*. Available at: https://www.joneslanglasalle.com.cn/en/newsroom/beijing-shanghai-and-shenzhen-listed-in-global-top-15-innovative-markets (Accessed: 1 November 2024).

Johnson, B. (2008) 'Cities, systems of innovation and economic development', *Innovation*, 10(2–3), pp. 146–155. Available at: https://doi.org/10.5172/impp.453.10.2-3.146.

Li, Y. and Du, R. (2022) 'Polycentric urban structure and innovation: evidence from a panel of Chinese cities', *Regional Studies*, 56(1), pp. 113–127. Available at: https://doi.org/10.1080/00343404.2021.1886274.

Liu, J., Meng, B. and Shi, C. (2023) 'A multi-activity view of intra-urban travel networks: A case study of Beijing', *Cities*, 143, p. 104634. Available at: https://doi.org/10.1016/j.cities.2023.104634.

Lotfata, A. and Georganos, S. (2023) 'Spatial machine learning for predicting physical inactivity prevalence from socioecological determinants in Chicago, Illinois, USA', *Journal of Geographical Systems* [Preprint]. Available at: https://doi.org/10.1007/s10109-023-00415-y.

Lotfata, A., Grekousis, G. and Wang, R. (2023) 'Using geographical random forest models to explore spatial patterns in the neighborhood determinants of hypertension prevalence across chicago, illinois, USA', *Environment and Planning B Urban Analytics and City Science* [Preprint]. Available at: https://doi.org/10.1177/23998083231153401.

Oh, D.-S. *et al.* (2016) 'Innovation ecosystems: A critical examination', *Technovation*, 54, pp. 1–6. Available at: https://doi.org/10.1016/j.technovation.2016.02.004.

OpenAI (2023) 'ChatGPT-4'. Available at: https://chat.openai.com.

People's Government of Beijing Municipality (2018) *Beijing's Xicheng District implements a block division plan to reshape the ecology of the old city*. Available at: https://www.beijing.gov.cn/renwen//jrbj/csjz/201808/t20180822_1874277.html (Accessed: 20 September 2024).

Shi, Q. *et al.* (2023) 'UGS-1m: fine-grained urban green space mapping of 31 major cities in China based on the deep learning framework', *Earth System Science Data*, 15(2), pp. 555–577. Available at: https://doi.org/10.5194/essd-15-555-2023.

Taecharungroj, V. and Millington, S. (2022) 'Amenity mix of innovation districts', *Journal of Place Management and Development*, 16(1), pp. 125–143. Available at: https://doi.org/10.1108/JPMD-05-2022-0042.

Tan, Y., Qian, Q. and Chen, X. (2023) 'Innovation District Space and Element Identification Framework: Empirical Research from Shenzhen, China', *Journal of Urban Planning and Development*, 149(1), p. 05022041. Available at: https://doi.org/10.1061/(ASCE)UP.1943-5444.0000884.

Ullman, E.L. (1954) 'Amenities as a Factor in Regional Growth', *Geographical Review*, 44(1), p. 119. Available at: https://doi.org/10.2307/211789.

Wang, S. *et al.* (2024) 'Geographically weighted machine learning for modeling spatial heterogeneity in traffic crash frequency and determinants in US', *Accident Analysis & Prevention*, 199, p. 107528. Available at: https://doi.org/10.1016/j.aap.2024.107528.

Yigitcanlar, T., Adu-McVie, R. and Erol, I. (2020) 'How can contemporary innovation districts be classified? A systematic review of the literature', *Land Use Policy*, 95, p. 104595. Available at: https://doi.org/10.1016/j.landusepol.2020.104595.

Yuan, X. *et al.* (2021) 'Innovation Districts in Beijing: Evolution, Distribution, and Development Mechanisms', in L. Bian, Y. Tang, and Z. Shen (eds) *Chinese Urban Planning and Construction: From Historical Wisdom to Modern Miracles*. Cham: Springer International Publishing, pp. 275–299. Available at: https://doi.org/10.1007/978-3-030-65562-4_13.

Zhang, C., Xiong, M. and Wei, X. (2022) 'Influence of Accessibility to Urban Service Amenities on Housing Prices: Evidence from Beijing', *Journal of Urban Planning and Development*, 148(1), p. 05021063. Available at: https://doi.org/10.1061/(ASCE)UP.1943-5444.0000795.

Zhou, X. and Yang, S. (2024) 'Distribution Characteristics and Influencing Factors of High-tech Enterprises in Beijing', *Economic geography*, 44(8), pp. 126–136.