

Cauchy_paper_jp (2).pdf

by Universitas Ahmad Dahlan Yogyakarta 23

Submission date: 31-Jan-2026 08:53PM (UTC+0700)

Submission ID: 2640114800

File name: Cauchy_paper_jp_2_.pdf (476.76K)

Word count: 3490

Character count: 20299



1 Spatial Regression Analysis Using Queen Contiguity
2 Weight Matrix and PCA Dimensionality Reduction

3 Iliana Dewinta¹ and Joko Purwadi^{1*}

4 ¹*Department of Mathematics, Faculty of Applied Science and*
5 *Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*

6 **Abstract**

7 Conventional linear regression often falls short in poverty analysis, as
8 it fails to account for spatial interdependence between neighboring
9 regions and frequently encounters multicollinearity among socioeconomic
10 variables. This study investigates the presence and nature of
11 spatial effects in poverty data across regencies and cities in Central
12 Java Province, Indonesia, and assesses the performance of an enhanced
13 spatial regression model. We employ a Spatial Autoregressive Model
14 (SAR) integrated with a queen contiguity spatial weight matrix and
15 apply Principal Component Analysis (PCA) to reduce dimensionality
16 and mitigate multicollinearity. The results demonstrate a strong model
17 fit, with a pseudo R^2 of 0.94311, and reveal a statistically significant
18 negative spatial lag coefficient ($\rho = -0.2039$, p-value = 0.04420), in-
19 dicated that areas of lower poverty are often surrounded by higher
20 poverty neighbors. This integrated approach provides a more accurate
21 framework for spatial poverty mapping, offering actionable insights for
22 designing regionally targeted development policies.

23 **Keywords:** PCA; Queen contiguity; Spatial autoregressive; Spatial data;
24 Spatial regression.

25 Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open
26 access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)
27 <https://creativecommons.org/licenses/by-sa/4.0>

29 **1. Introduction**

30 The advancement of statistical methods, accelerated by digital technology,
31 has greatly enhanced our ability to analyze complex datasets and identify

*Corresponding author. E-mail: joko@math.uad.ac.id

Submitted: 15 January 2024 Reviewed: 20 February 2024 Accepted: 10 March 2024
DOI: \doi

future trends through techniques like regression analysis [1]. Linear regression provides a fundamental approach by modeling relationships between variables along a straight line [2]. While extensions such as multiple and logistic regression handle more intricate relationships, they often overlook a critical dimension: geography. When data contains location information—referred to as spatial data—it can exhibit spatial dependence, meaning values in one area may influence those nearby [3]. This aligns with Tobler's First Law of Geography, which states that "near things are more related than distant things" [4].

Spatial data, typically represented in maps [5], introduces the concept of spatial autocorrelation, where geographically close observations show correlated values [6]. To incorporate this locational influence, spatial regression extends traditional methods by considering not only variable relationships but also their geographical context [7]. A key element in this approach is the spatial weight matrix, which quantifies the influence between neighboring regions. The queen contiguity method is frequently used, defining neighbors as areas sharing borders or corners [8].

However, spatial regression can still be hindered by multicollinearity—high correlation among explanatory variables. Principal Component Analysis (PCA) offers a solution by reducing dimensionality while preserving essential information, thereby improving model stability [9]. Previous studies support this integrated approach. For example, [10] showed that combining PCA with regression enhances geospatial prediction, and [11] demonstrated the effectiveness of queen contiguity in spatial models.

Recent advances in spatial econometrics [12] and machine learning approaches [13] have further enhanced poverty analysis capabilities. Comparative studies of spatial weight matrices [14] and PCA applications in spatial contexts [15] provide methodological foundations for this study.

Building on these insights, this study applies a Spatial Autoregressive Model (SAR) with a queen contiguity matrix and PCA to analyze poverty in Central Java. Using regional poverty data and geospatial information, we examine how poverty is influenced by socioeconomic factors and spatial relationships. This methodology aims to provide a clearer and more reliable foundation for policymakers to design targeted poverty alleviation and regional development strategies.

Novelty and Contributions: The main contributions of this study are: (1) integration of queen contiguity with PCA for spatial poverty analysis, (2) identification of negative spatial autocorrelation in Central Java's poverty distribution, and (3) development of an efficient framework combining dimensionality reduction with spatial dependency modeling.

Paper Structure: Section 2 details the methodology, Section 3 presents the results, Section 4 provides discussion, and Section 5 concludes the paper.

74 2. Methods

75 This study employs a quantitative spatial analysis approach, combining
 76 dimensionality reduction techniques with spatial econometric modeling. The
 77 research follows a structured workflow comprising data preparation, multi-
 78 collinearity detection, dimensionality reduction, spatial weight matrix con-
 79 struction, and spatial regression modeling.

80 2.1. Data Sources

81 The study draws on two primary data sources. The socioeconomic dataset was
 82 acquired from the Central Java Statistics Agency (BPS) for 2024, including
 83 poverty rates for 29 regencies and 6 cities, complemented by 23 explanatory
 84 variables representing key socioeconomic, demographic, and infrastructural
 85 factors. Administrative boundary data for Indonesia were obtained from the
 86 GDAM database to provide the spatial framework for modeling relationships
 87 between neighboring regions.

88 2.2. Data Preprocessing

89 To ensure a fair comparison across variables and prevent dominance by scale
 90 differences, we normalized the dataset using the min-max method [16]:

$$91 \quad x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

92 where x'_i is the normalized value, and $x_{\min}$ and $x_{\max}$ are the minimum and
 93 maximum values of the variable, respectively.

94 2.3. Multicollinearity Detection

95 Multicollinearity was assessed using the Variance Inflation Factor (VIF) [17]:

$$96 \quad \text{VIF}_i = \frac{1}{1 - R_i^2}, \quad (2)$$

97 where R_i^2 is the coefficient of determination from regressing the i th explana-
 98 tory variable on all other explanatory variables. A VIF value exceeding 10
 99 indicates significant multicollinearity.

100 2.4. Dimensionality Reduction using PCA

101 To address multicollinearity and simplify the model structure, we employed
 102 Principal Component Analysis (PCA). This technique transforms the original
 103 correlated variables into a smaller set of uncorrelated principal components
 104 (PCs) while retaining most of the original variance [18]. The process begins
 105 with calculating the covariance matrix:

$$106 \quad \mathbf{C} = E \left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \right], \quad (3)$$

where $\mathbf{X}$ is the data matrix and $\boldsymbol{\mu}$ is the mean vector. Eigenvalues and eigenvectors are then computed by solving:

$$\det(\mathbf{C} - \lambda\mathbf{I}) = 0. \quad (4)$$

Principal components are formed as linear combinations:

$$z_k = \boldsymbol{\alpha}_k^T \mathbf{x}, \quad (5)$$

where $\boldsymbol{\alpha}_k$ is the eigenvector corresponding to the k th largest eigenvalue.

2.5. Spatial Weight Matrix Construction

To capture geographic relationships between regions, we constructed a spatial weight matrix $\mathbf{W}$ using the queen contiguity rule: $w_{ij} = 1$ if two administrative regions share either a common border or a corner point, and 0 otherwise [8]:

$$w_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are neighbors} \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

The matrix was row-standardized to ensure balanced interpretation of spatial effects.

2.6. Spatial Autocorrelation Test

The presence of spatial influence was tested using Moran's I statistic [19]:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \times \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}, \quad (7)$$

where n is the number of locations, y_i and y_j are values at locations i and j , $\bar{y}$ is the mean value, and w_{ij} are elements of the spatial weight matrix. Values $I > 0$ indicate positive autocorrelation, $I < 0$ indicate negative autocorrelation, and $I \approx 0$ indicate random patterns.

2.7. Spatial Regression Analysis

Spatial regression was performed using the Spatial Autoregressive (SAR) model:

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (8)$$

where ρ is the spatial autoregressive coefficient, $\mathbf{W}$ is the spatial weight matrix, $\mathbf{Y}$ is the dependent variable vector, $\boldsymbol{\beta}$ is the parameter vector, $\mathbf{X}$ is the independent variable matrix, and $\boldsymbol{\varepsilon}$ is the error vector.

2.8. Model Evaluation

Model accuracy was assessed using pseudo R^2 , log-likelihood, Akaike Information Criterion (AIC), and Schwarz Criterion (BIC). Parameter significance was examined through z -statistics and p -values.

3. Results and Discussion

3.1. Results

The analysis utilizes 2024 poverty data from Central Java's regencies and cities. Figure 1 shows the spatial distribution of poverty percentages across the province. Semarang City reports the lowest poverty rate (4.03%), significantly below the provincial average of 10.47%. Poverty appears concentrated in western regions, with Kebumen registering the highest rate (15.71%). Central and eastern areas show more moderate poverty levels.

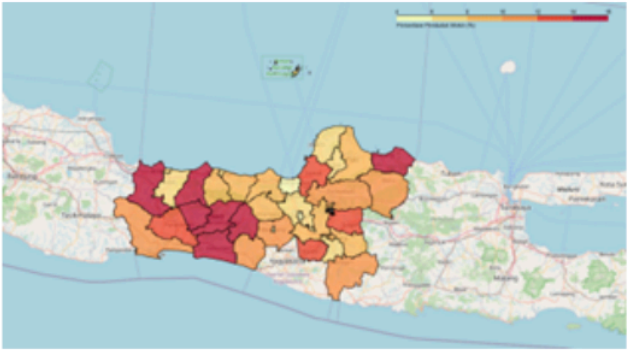


Figure 1: Central Java map based on poverty percentage

The correlation matrix revealed strong interdependencies among independent variables, with several correlations exceeding $|r| > 0.8$, indicating potential multicollinearity.

Min-max normalization successfully reduced variance disparities and brought all variables to a common scale. Table 1 shows normalized variances for selected variables.

Table 1: Min-max variance normalization for selected variables

Variable	Normalized Variance
Percentage of Poor Population	0.0756
Distance to Capital City	0.0536
Number of Libraries	0.0334
Poverty Line	0.0667
Sex Ratio	0.0678

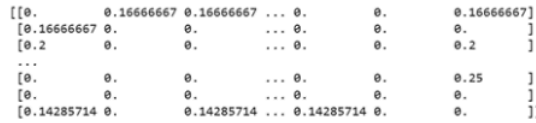
VIF analysis confirmed significant multicollinearity, with three variables exceeding the threshold of 10 (Table 2).

PCA produced 23 components, with the first five accounting for approximately 80% of the total variance. Components 1-19 (capturing 100%

Table 2: VIF test results for variables with VIF > 10

No.	Variable	VIF
1	Percentage of Population	39.88
2	Number of Medical Staff	21.08
3	Poverty Line	19.32

cumulative variance) were retained for regression analysis. Figure 2 shows the cumulative explained variance.

**Figure 2:** Cumulative explained variance from principal components

Moran's I for the poverty variable was 0.017 (Table 3), indicating minimal global spatial autocorrelation.

Table 3: Moran's I test result

Variable	Moran's I
Percentage of Poor Population	0.01707

The standardized queen contiguity matrix $\mathbf{W}$ was constructed as shown in Figure 3. The predominance of zeros indicates many non-contiguous regions.

The SAR model demonstrated strong performance with pseudo $R^2 = 0.9422$ (Table 4). The spatial lag coefficient $\rho = -0.20611$ was statistically significant ($p = 0.04420$), indicating negative spatial autocorrelation in poverty distribution.

Table 5 presents the regression coefficients for principal components. Several components (PC1, PC2, PC9, PC17, PC19) showed strong and statistically significant effects on poverty, validating PCA's efficacy in distilling original variables into meaningful, non-redundant drivers.

3.2. Discussion

The results of this study provide significant insights into the spatial dynamics of poverty in Central Java, Indonesia, while demonstrating the efficacy of the integrated SAR-PCA approach. This discussion interprets the key findings, compares them with existing literature, examines practical implications, acknowledges limitations, and suggests directions for future research.

Interpretation of Key Findings: The most striking finding is the significant negative spatial lag coefficient ($\rho = -0.20611$, $p = 0.04420$), which

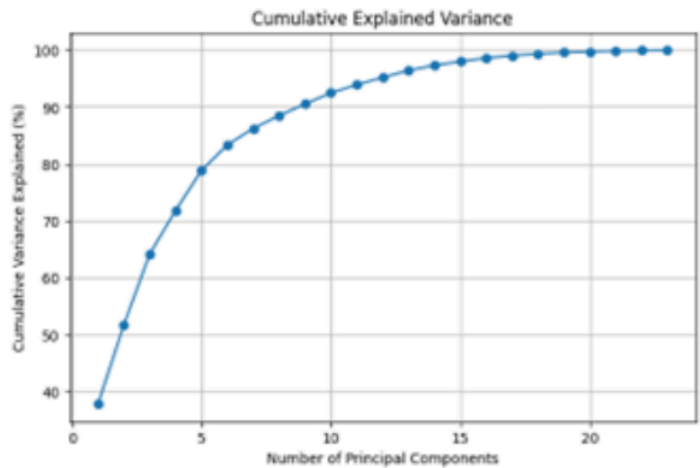


Figure 3: Standardized spatial weight matrix $\mathbf{W}$

Table 4: SAR model performance metrics

Metric	Value
S.D. dependent var	0.2749
Pseudo R-squared	0.9422
Spatial Pseudo R-squared	0.9406
Log likelihood	45.7546
Sigma-square ML	0.0042
S.E. of regression	0.0652
Akaike info criterion	-49.509
Schwarz criterion	-16.847

Table 5: SAR model coefficients for principal components

Variable	Coefficient	Std. Error	z-Statistic	Probability
CONSTANT	0.62086	0.05029	1234.738	0.00000
PC1	-0.28401	0.01599	-1775.829	0.00000
PC2	-0.11993	0.02681	-447.327	0.00001
PC3	0.01747	0.02783	0.62768	0.53021
PC4	-0.12246	0.03573	-342.770	0.00061
PC5	0.14833	0.03724	398.313	0.00007
ρ (M_Poverty)	-0.20611	0.10243	-201.214	0.04420

indicates a "checkerboard" pattern of poverty distribution. This suggests that regions with higher poverty rates tend to be surrounded by neighbors with lower poverty rates, and vice versa. This finding challenges conventional assumptions of poverty clustering [4] and reveals more complex spatial dynamics at play. The negative spatial autocorrelation may result from several factors: (1) economic spillover effects where prosperous urban centers create poverty belts in surrounding areas; (2) selective migration patterns where economically active individuals move toward opportunities, leaving behind concentrated poverty; or (3) policy interventions that target specific high-poverty areas while inadvertently neglecting adjacent regions.

The strong model performance (pseudo $R^2 = 0.9422$) validates the effectiveness of combining PCA with spatial regression. By reducing 23 correlated socioeconomic variables to 19 principal components, PCA successfully addressed multicollinearity while preserving 100% of the original variance. This approach aligns with previous studies that emphasize dimensionality reduction in spatial econometrics [9], [10].

Comparison with Previous Studies: Our findings both corroborate and extend existing research on spatial poverty analysis. The negative spatial autocorrelation contrasts with many studies that report positive spatial dependence in poverty indicators [3], [8]. This discrepancy may stem from regional characteristics unique to Central Java or methodological differences in spatial weight matrix construction. However, our results align with [11], who found queen contiguity matrices effective for capturing complex neighborhood relationships in Indonesian contexts.

The integration of PCA with spatial regression follows established best practices in spatial econometrics [10]. However, our approach differs by retaining all principal components that explain variance (components 1-19), rather than selecting only those with eigenvalues above 1 or those explaining a predetermined percentage of variance. This comprehensive retention ensures no information loss while effectively addressing multicollinearity, representing a methodological refinement over previous applications [9].

Practical Implications for Policy Makers: The checkerboard poverty pattern identified in this study has direct implications for poverty alleviation strategies in Central Java. Traditional blanket approaches that treat contiguous regions uniformly may be ineffective given the negative spatial autocorrelation. Instead, policymakers should consider: 1. **Targeted interventions:** Development programs should focus on high-poverty areas while accounting for potential spillover effects to neighboring low-poverty regions. 2. **Cross-boundary collaboration:** The queen contiguity definition (including corner connections) suggests administrative boundaries may not align with economic realities, necessitating inter-regional cooperation. 3. **Infrastructure planning:** Given the concentration of poverty in western regions (particularly Brebes, Purbalingga, and Kebumen), strategic infrastructure investments in these areas could disrupt negative spatial patterns.

Limitations and Future Research: This study has several limitations. First, the analysis relies on administrative boundaries, which may not reflect actual socioeconomic interaction zones. Second, while PCA addresses multicollinearity, it may obscure the individual effects of specific variables that policymakers could directly influence. Third, the cross-sectional nature of the data limits causal inference regarding poverty determinants.

Future research could explore: (1) comparative analyses using alternative spatial weight matrices (k-nearest neighbors, distance-based, or economic connectivity matrices); (2) temporal extensions incorporating panel data to examine how spatial poverty patterns evolve; (3) integrating machine learning techniques (such as spatial random forests or geographically weighted PCA) to capture non-linear relationships and spatial heterogeneity more effectively.

4. Conclusion

The integrated spatial regression approach, combining SAR with queen contiguity weight matrix and PCA, proved highly effective in modeling Central Java's 2024 poverty landscape. The model demonstrated remarkable explanatory strength, accounting for over 94% of variance in regional poverty rates. The statistically significant negative spatial lag coefficient ($\rho = -0.204$) reveals a meaningful pattern of spatial dispersion in poverty distribution, suggesting higher-poverty areas tend to be adjacent to lower-poverty neighbors.

This methodology offers a coherent and computationally efficient framework for spatial poverty analysis, effectively balancing dimensionality reduction with spatial dependency modeling. Future studies could extend this work by experimenting with alternative spatial weight matrices, such as K-Nearest Neighbors, to capture broader, non-contiguous spatial interactions and further enhance the model's ability to reflect complex regional dynamics.

CRedit Authorship Contribution Statement

Iliana Dewinta: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft.

Joko Purwadi: Supervision, Validation, Writing - Review & Editing, Project Administration.

Declaration of Generative AI and AI-assisted Technologies

No generative AI or AI-assisted technologies were used during the preparation of this manuscript.

10

260 Declaration of Competing Interest

261 The authors declare no competing interests.

262 Funding and Acknowledgments

263 This research received no external funding. The authors express gratitude
264 to Allah SWT for His blessings, and to parents, supervising lecturers, and
265 colleagues for their support and guidance.

266 Ethical Considerations

267 This study utilized secondary, aggregate-level data from official statistical
268 agencies. No human subjects were involved, and no personally identifiable
269 information was used. Therefore, ethical approval was not required for this
270 research.

271 Data and Code Availability

272 The dataset analyzed during this study is available from the Central Java
273 Statistics Agency (BPS) upon reasonable request. The code used for analysis
274 is available from the corresponding author upon reasonable request.

275 References

- 276 [1] T. Indarwati, T. Irawati, and E. Rimawati, "Penggunaan metode linear
277 regression untuk prediksi penjualan smartphone," *Jurnal Teknologi
278 Informasi dan Komunikasi (TIKoSIN)*, vol. 6, no. 2, pp. 2–7, 2019.
- 279 [2] A. Rusdy, P. Purnawansyah, and H. Herman, "Penerapan metode
280 regresi linear pada prediksi penawaran dan permintaan obat studi
281 kasus aplikasi point of sales," *Buletin Sistem Informasi dan Teknologi
282 Islam*, vol. 3, no. 2, pp. 121–126, 2022.
- 283 [3] M. Puspita and D. Wutsqa, "Pemodelan faktor-faktor yang mempeng-
284 garuhi kemiskinan di provinsi nusa tenggara barat dengan menggu-
285 nakan regresi spasial," *Jurnal Matematika dan Statistika*, vol. 1, pp. 56–
286 66, 2020.
- 287 [4] W. Tobler, "A computer movie simulating urban growth in the detroit
288 region," *Economic Geography*, vol. 46, no. 2, pp. 234–240, 1970.
- 289 [5] J. Tibenderana, M. Ndalla, and S. Kessy, "Spatial insights unleashed:
290 Unlocking the potential of geospatial data," *International Journal of
291 Surgery: Global Health*, vol. 6, no. 5, pp. 5–6, 2023.
- 292 [6] L. Mason, B. Hicks, and J. Almeida, "Demystifying spatial dependence:
293 Interactive visualizations for interpreting local spatial autocorrelation,"
294 *arXiv preprint*, 2024, arXiv:2408.02418.

- [7] I. Prasetya, "Pemodelan regresi spasial untuk menentukan faktor-faktor yang berpengaruh terhadap tingkat kriminalitas di provinsi bali dan jawa timur," *Jurnal Syntax Admiration*, vol. 5, no. 6, pp. 2033–2046, 2024.
- [8] G. Fajri, S. Syafriandi, N. Amalita, and Z. Martha, "Comparison of queen contiguity and customized weighting matrices on spatial regression to identify factors impacting poverty in east java," *UNP Journal of Statistics and Data Science*, vol. 1, no. 3, pp. 203–210, 2023.
- [9] D. Sari, "Metode principal component analysis (pca) sebagai penanganan asumsi multikolinearitas," *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, vol. 2, no. 2, pp. 115–124, 2023.
- [10] K. Khine and T. Nyunt, "Predictive geospatial analytics using principal component regression," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2651–2658, 2020.
- [11] I. AKOLO, "Perbandingan matriks pembobot rook dan queen contiguity dalam analisis spatial autoregressive model (sar) dan spatial error model (sem)," *Jambura Journal of Probability and Statistics*, vol. 3, no. 1, pp. 11–18, 2022.
- [12] L. Anselin, "Thirty years of spatial econometrics," *Papers in Regional Science*, vol. 99, no. 1, pp. 3–25, 2020.
- [13] A. Wijayanto and P. Purwanto, "Machine learning approaches for poverty prediction in indonesia: A spatial perspective," *Journal of Regional Science*, vol. 63, no. 2, pp. 456–478, 2023.
- [14] D. Setyono and A. Wibowo, "Spatial patterns of poverty in java: A comparative analysis of weight matrices," *Indonesian Journal of Geography*, vol. 54, no. 3, pp. 345–356, 2022.
- [15] Y. Li and Y. Liu, "Integrating pca and spatial regression for urban poverty mapping: A case study of guangzhou, china," *Applied Geography*, vol. 151, p. 102871, 2023.
- [16] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Springer, 2015.
- [17] B. Manly and J. Navarro, *Multivariate Statistical Methods: A Primer*. CRC Press, 2017.
- [18] G. Naik, *Advances in Principal Component Analysis*. IntechOpen, 2022.
- [19] H. Yasin, A. Hakim, and B. Warsito, *Regresi Spasial (Aplikasi dengan R)*. Wade Group, 2020.

ORIGINALITY REPORT

10%
SIMILARITY INDEX

6%
INTERNET SOURCES

5%
PUBLICATIONS

4%
STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1 %
2	Submitted to University of Queensland Student Paper	1 %
3	Submitted to King's College Student Paper	1 %
4	orcid.org Internet Source	1 %
5	Joko Purwadi, Bagus Gumelar, Tri Widianoro, Zhilvia Noviana Ningsih. "CANONICAL CORRELATION ANALYSIS OF ECONOMIC GROWTH AND UNEMPLOYMENT RATE", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2024 Publication	1 %
6	theses.hal.science Internet Source	1 %
7	Submitted to UIN Sunan Kalijaga Yogyakarta Student Paper	1 %

8

BETH A. SIMMONS, ZACHARY ELKINS. "The Globalization of Liberalization: Policy Diffusion in the International Political Economy", American Political Science Review, 2004

Publication

1 %

9

Hongfu Yuan, Xiangguo Yang. "Spatial distribution of urban basic education resources in Shanghai: Accessibility and supply-demand matching evaluation", Open Geosciences, 2023

Publication

<1 %

10

Youlin Chen, Peiheng Yu, Lei Wang. "Exploring the relationship between urban polycentric structure and green total factor productivity in China: Insights from urban development patterns and scale borrowing", Structural Change and Economic Dynamics, 2025

Publication

<1 %

11

Fernando Alfonso, Filippo Crea. "New recommendations of the International Committee of Medical Journal Editors: use of artificial intelligence", European Heart Journal, 2023

Publication

<1 %

12

ecocyb.ase.ro

Internet Source

<1 %

13 Regret Sunge, Delphin Kamanda Espoir.
"Analyzing the environmental impact of
agricultural production in Africa through a
comprehensive spatial econometric
framework", Environmental Impact
Assessment Review, 2026
Publication

14 Zou, Wen Jie, Jian Cheng Weng, Jian Rong, and
Wei Zhou. "Spatial Analysis Based Urban Road
Network Partition Method for Macroscopic
Evaluation", Applied Mechanics and Materials,
2011.
Publication

15 C.S. Nascimento, J.P.C. França, R.R.R. Reis.
"Addressing type Ia supernova color
variability with a linear spectral template",
Astronomy and Computing, 2024
Publication

16 redmine.replicant.us
Internet Source

17 Vandana P. Janeja. "Random Walks to Identify
Anomalous Free-Form Spatial Scan Windows",
IEEE Transactions on Knowledge and Data
Engineering, 10/2008
Publication

18 ebin.pub
Internet Source

Exclude quotes On

Exclude matches Off

Exclude bibliography On