Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Composite Evaluation of High-Speed Rail Station Infrastructure in China: An AHP–Entropy Weighting and Sensitivity Analysis

Renjie Jin

ffgxfvy5562@outlook.com

China University of Petroleum-Beijing at Karamay https://orcid.org/0009-0004-3436-2453

Research Article

Keywords: high-speed rail, infrastructure appraisal, AHP, entropy weighting, composite index, sensitivity analysis, China

Posted Date: July 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10521795/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Infrastructure appraisal requires transparent integration of economic returns, spatial coverage, service use, and capital cost. We conducted a secondary methodological analysis of a station-level dataset assembled from public statistical releases, city reports, satellite-derived catchment estimates, and database exports. Eighteen Chinese high-speed rail stations with complete data were evaluated using four indicators: economic benefit, catchment area, annual passenger volume, and construction cost. Analytic Hierarchy Process (AHP) weights were combined equally with entropy weights, after min–max normalization and reverse scaling of cost. The resulting weights were 0.435, 0.165, 0.290, and 0.110, respectively. Composite scores ranged from 15.03 to 70.31 (mean 38.40; median 33.03). Guangzhou South ranked first, followed by Shanghai Hongqiao and Xi'an North; Lanzhou West and Yangshuo had the lowest relative scores. In 10,000 weight-perturbation analyses in which each weight varied by ± 20% before renormalization, the median rank correlation with the baseline ranking was 0.996 and the fifth percentile was 0.990; top-five membership was preserved in at least 95% of simulations. Beijing indicators for 2002–2014 were examined descriptively as temporal demand context, not as a validated forecasting exercise. The framework provides an auditable relative comparison, but it does not measure causal returns, social welfare, or the absolute necessity of a station. Policy application requires source-level provenance, uncertainty in indicator construction, out-of-sample validation, and stakeholder-defined decision thresholds.

1. INTRODUCTION

High-speed rail can improve accessibility and reshape regional economic interactions, but infrastructure benefits depend on network position, demand, complementary investment, and land-use conditions [1]. Large transport projects may also redistribute activity rather than create uniform gains, so appraisal should distinguish network-wide strategic value from station-level operating evidence [2]. In China, rapid network expansion has increased the need for reproducible methods that compare service use and spatial coverage with financial commitment without reducing the decision to a single unexamined indicator.

Multi-criteria decision analysis offers a structured way to combine heterogeneous dimensions. The Analytic Hierarchy Process (AHP) represents expert judgments through pairwise comparisons [3], and integrated AHP applications are common when both qualitative and quantitative evidence matter [4]. Entropy weighting uses dispersion in observed data to derive an objective information weight [5]. Neither method is neutral: AHP depends on the judgment matrix, while entropy rewards variation even when variation is not synonymous with policy importance.

The source statistical-modeling report combined AHP and entropy weights to assess station infrastructure and explored Beijing time-series indicators as a broader demand context. The present study converts that work into a focused research article, recalculates the documented station scores, adds a prespecified weight-sensitivity analysis, and narrows interpretation to relative comparison. Our

objectives were to quantify score distributions, evaluate ranking stability, describe indicator trade-offs, and specify the evidence required before the scores could inform investment decisions.

2. MATERIALS AND METHODS

2.1 Study design and data provenance

We performed a secondary computational analysis of aggregate station and city indicators. The supplied station workbook contained 20 named stations; two records lacked passenger-volume or catchment-area values, leaving 18 complete cases. Indicators had been assembled from public city and statistical reports, media reports, and satellite-based area calculations. Beijing time-series exports identified the National Bureau of Statistics as the original source for population and macroeconomic indicators [11] and were accessed through RESSET demonstration-account files [12]. No individual-level or identifiable data were used.

The four station-level criteria were total economic benefit (100 million CNY), catchment area (km²), annual passengers (10,000 persons), and construction cost (100 million CNY). Economic benefit, area, and passengers were benefit criteria; construction cost was a cost criterion. Because the records combined years near 2019 and source definitions were not fully harmonized, the analysis was treated as a methodological benchmark rather than a contemporaneous official ranking.

2.2 AHP and entropy weights

The source AHP comparison matrix yielded weights of 0.49, 0.16, 0.25, and 0.10 for economic benefit, catchment area, passengers, and cost. AHP consistency should be evaluated through the consistency ratio in any operational implementation. Entropy weights were 0.38, 0.17, 0.33, and 0.12. The entropy of criterion j was calculated from normalized proportions p_{ij}:

$$E_j = -(1/\ln n) \sum_i p_{ij} \ln(p_{ij}) \quad (1)$$

AHP and entropy weights were averaged with an equal mixing coefficient of 0.5. This produced combined weights of 0.435, 0.165, 0.290, and 0.110 (Fig. 1). Equal mixing was retained to reproduce the source analysis; it is a modeling choice, not an empirically established optimum. The general logic is consistent with integrated AHP–entropy applications in infrastructure-risk assessment [7].

2.3 Normalization and composite score

Benefit criteria were min–max normalized as $(x_{ij} - \min x_j)/(\max x_j - \min x_j)$. Construction cost was reverse-scaled as $(\max x_j - x_{ij})/(\max x_j - \min x_j)$. Min–max scaling preserves relative position but is sensitive to sample extremes; the resulting scores are therefore sample dependent, as with other compensatory multi-criteria indices [6].

$$\text{Score}_i = 100 \times \sum_j w_j z_{ij} \quad (2)$$

Scores were grouped using the source thresholds: below 20, 20 to below 30, 30 to below 50, and at least 50. These labels describe relative positions in this benchmark and are not validated cutoffs for project approval, closure, or future investment. To verify reproducibility, calculated scores were compared with the source table after rounding to two decimals.

2.4 Weight-sensitivity analysis

We generated 10,000 deterministic-seed perturbation scenarios. Each baseline weight was independently sampled from 80% to 120% of its original value and the resulting vector was renormalized to sum to one. We calculated the rank correlation with the baseline order and top-five overlap. The ± 20% interval was chosen as a transparent robustness stress test, not as a probabilistic confidence distribution.

2.5 Temporal demand context

Beijing resident population, per-capita GDP, per-capita road area, population density, and railway operating mileage were aligned for the common 2002–2014 period. Each series was standardized to a z-score for visual comparison. Following time-series reporting principles [9], the plot was used descriptively because 13 annual observations are insufficient to support strong claims for a high-capacity forecasting model. Although LSTM networks can represent long-term dependence [10], no BP–LSTM performance claim was retained without reproducible holdout errors and baseline comparisons.

3. RESULTS

3.1 Score distribution and ranking

All recalculated scores agreed with the source table within 0.01 points. Scores ranged from 15.03 to 70.31, with a mean of 38.40 and median of 33.03. Five stations scored at least 50, five scored from 30 to below 50, six scored from 20 to below 30, and two scored below 20. The full ranking is shown in Fig. 2.

Guangzhou South ranked first (70.31), followed by Shanghai Hongqiao (62.76), Xi’an North (60.41), Beijing Fengtai (59.52), and Hangzhou East (53.29). Lanzhou West (15.03) and Yangshuo (16.12) had the lowest scores. These differences reflect the chosen normalization and compensatory weights; they do not establish that lower-scoring stations lack wider network, equity, emergency, or strategic value.

3.2 Indicator profiles

The normalized profiles in Fig. 3 show that high scores arose through different combinations. Guangzhou South combined the highest passenger volume with strong economic benefit, whereas Shanghai Hongqiao combined high economic benefit and catchment coverage but was penalized by the maximum construction cost. Xi’an North received the maximum catchment-area value and high passenger volume. Low-scoring records generally lacked a compensating advantage on heavily weighted benefit criteria.

3.3 Sensitivity to weight perturbation

The median rank correlation across perturbations was 0.996; the fifth percentile was 0.990. Top-five membership was unchanged in at least 95% of simulations. This indicates that the leading group was stable to moderate variation around the selected weights. Stability does not validate indicator definitions or data provenance: a ranking can be mathematically robust while remaining incomplete as a policy appraisal.

3.4 Beijing temporal context

From 2002 to 2014, resident population and per-capita GDP rose strongly, while railway operating mileage increased more gradually and with periods of limited change. Population density also increased, whereas per-capita road area fluctuated rather than following a monotonic trajectory (Fig. 4). The divergence demonstrates why infrastructure demand should not be inferred from a single macroeconomic trend.

3.5 Complete-case accounting and reproducibility

The supplied station workbook contained 20 named records. Taiyuan South lacked both passenger volume and catchment area, and Nanjing South lacked catchment area; both were excluded before normalization. No values were statistically imputed. Because min–max bounds were calculated from the 18 complete cases, adding the two stations after future data retrieval could change every normalized value and score even when the original observations remain unchanged. The present ranking therefore applies only to the documented complete-case set.

Recalculation from the four raw indicators reproduced all published station scores after rounding, with an average absolute difference below 0.01 points. This arithmetic agreement verifies the stated normalization and weighting procedure. It does not verify the accuracy of the underlying public-report values, satellite-derived catchment boundaries, or economic-benefit construction. Those elements require a source-level audit and should be released as a machine-readable provenance table in any subsequent version.

4. DISCUSSION

4.1 Principal findings

The combined AHP–entropy framework reproduced an interpretable station ranking and remained stable under moderate weight perturbation. Economic benefit received the largest combined weight, followed by passenger volume, catchment area, and cost efficiency. The method therefore emphasized realized or estimated utilization benefits more than financial economy. This prioritization explains why several high-cost stations still ranked highly when passenger and economic indicators were strong.

The sensitivity analysis is an important addition because a point ranking alone can conceal dependence on subjective weights. Here, stability of the top group suggests that small changes to the mixing assumptions would not reverse the principal result. Lower and middle ranks may still be sensitive to data updates, missing-value treatment, and source-year harmonization.

4.2 Interpretation for infrastructure appraisal

The score is best used as a screening instrument that identifies cases for deeper review. A station with a lower relative score may serve redistribution, national connectivity, disaster redundancy, or long-term development objectives that are absent from the four indicators. Conversely, a high score does not demonstrate causal economic returns or cost effectiveness. Decision-makers should review the component indicators and uncertainty, not only the aggregate rank.

A complete appraisal should add lifecycle maintenance cost, travel-time savings, service frequency, network centrality, environmental effects, land-value changes, distributional access, and counterfactual demand. Regional differences in reporting and price years also require harmonization. Thresholds for action should be defined by responsible institutions and stakeholders rather than inferred from sample quantiles.

4.3 Strengths and limitations

Strengths include transparent equations, complete-case accounting, exact score reproduction, separation of benefit and cost directions, and a documented robustness stress test. The study also replaces an unsupported high-capacity forecasting claim with a descriptive temporal analysis aligned to the available common years.

Limitations are substantial. The station sample was purposive rather than probabilistic; two records were excluded because of missing data. Indicator years and definitions were heterogeneous, and some source values were compiled from media reports or satellite-derived calculations without a complete audit trail. Min–max normalization is sensitive to extremes, AHP judgments were not independently elicited, and equal AHP–entropy mixing was prespecified. The four criteria omit wider social and network functions. No external criterion, causal design, or out-of-sample validation was available.

4.4 Reproducibility and future validation

Before policy use, the derived station table, exact source URLs, access dates, geospatial procedures, comparison matrix, code, and software environment should be archived. Proprietary database exports should be handled under provider terms rather than redistributed without permission. Future studies should pre-register indicator definitions, use multiple years, incorporate uncertainty in measurements, and test whether scores predict independently observed outcomes such as ridership growth, travel-time gains, operating deficits, or accessibility improvements.

Robustness analysis should be expanded to alternative normalizations, missing-data methods, non-compensatory decision rules, and stakeholder-specific weight sets. External validation on stations not

used to define the scoring system would show whether the ranking framework generalizes beyond the original benchmark. These steps are necessary before assigning normative labels such as “necessary” or “unnecessary” to infrastructure projects.

4.5 Data-quality and governance requirements

Infrastructure rankings can influence public perceptions and resource allocation, so data governance is part of model validity. Each observation should be accompanied by a definition, reference year, geographic boundary, currency basis, source URL, retrieval date, and transformation history. Economic-benefit estimates should distinguish passenger revenue from wider economic contribution; catchment calculations should specify travel mode and threshold; passenger counts should distinguish boarding, alighting, and transfer flows; and construction cost should be converted to a common price year. Without these controls, apparent differences may reflect inconsistent measurement rather than infrastructure performance.

The publication of a composite score should also preserve disaggregated evidence. Users need to see when a high score is driven by one dominant indicator and when a low score results from missing dimensions. Governance safeguards should include independent review of the AHP matrix, sensitivity analyses for stakeholder-specific weights, versioned updates as new observations become available, and an explicit prohibition on using the score as the sole basis for project approval or closure. These practices would make the framework auditable while respecting the strategic and distributional functions that are not represented in the current four-criterion model.

4.6 Alternative specifications

Several alternative specifications should be examined before the score is used beyond this benchmark. Rank-based or winsorized normalization could reduce sensitivity to extreme values. Geometric aggregation could limit full compensation between very weak and very strong indicators, while outranking methods could identify pairwise dominance without assuming that all trade-offs are linear. A cost-effectiveness specification could also separate benefits from capital cost rather than combining them in one additive score. Comparing these approaches would show whether the leading stations are stable to the mathematical form of aggregation, not only to moderate changes in weights.

Uncertainty should be propagated from measurements as well as preferences. Passenger volume and economic contribution may have sampling, accounting, or temporal error; catchment areas depend on geographic assumptions; and construction costs may exclude land acquisition, financing, or subsequent expansion. Scenario intervals for each input could be propagated through the scoring system to produce rank acceptability rather than a single deterministic position. This would better distinguish clearly separated stations from those whose ordering is practically indeterminate.

4.7 Decision-use protocol

A responsible decision protocol could use three stages. First, the composite score would screen stations for additional review, with no automatic approval or rejection threshold. Second, analysts would examine

indicator-level evidence, uncertainty, and omitted strategic functions for each flagged station. Third, a multidisciplinary panel would consider distributional access, network redundancy, environmental impact, and long-term regional objectives before making recommendations. The score would therefore organize evidence rather than replace institutional judgment.

For longitudinal monitoring, the analysis should be updated on a fixed schedule using consistent definitions and price years. Changes in score should be decomposed into updated performance, revised weights, and changes to the comparison sample. Publishing this decomposition would prevent a station from appearing to improve or deteriorate solely because another station entered the dataset. Together, these procedures would convert the present cross-sectional demonstration into a more transparent infrastructure-governance tool.

5. CONCLUSION

This secondary methodological analysis provides a transparent and reproducible AHP–entropy evaluation of 18 Chinese high-speed rail stations. The combined score was dominated by economic benefit and passenger volume, and the leading five stations remained stable under moderate weight perturbation. The results support relative screening, not definitive judgments about infrastructure necessity or causal return. Better provenance, harmonized longitudinal data, broader social and network indicators, and external validation are required before the framework can inform investment decisions.

Declarations

Ethics approval and consent to participate

Not applicable. This secondary computational study used aggregate infrastructure and city-level indicators and involved no human participants, identifiable personal information, intervention, or animal research.

Consent for publication

Not applicable. No identifiable individual information is reported.

Competing interests

The author declares no competing financial or non-financial interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

R.J. conceived the study, assembled the source material, conducted the modeling and sensitivity analysis, interpreted the results, and wrote and revised the manuscript.

Data availability

The derived station-level analysis table and code are available from the corresponding author on reasonable request. RESSET demonstration-account exports and other third-party source files are subject to their original providers' access and redistribution terms and are not relicensed by this manuscript.

References

1. Givoni M (2006) Development and impact of the modern high-speed train: A review. Transp Reviews 26:593–611. https://doi.org/10.1080/01441640600589319
2. Vickerman R (1997) High-speed rail in Europe: Experience and issues for future development. Annals Reg Sci 31:21–38. https://doi.org/10.1007/s001680050037
3. Saaty TL (2008) Relative measurement and its generalization in decision making. RACSAM 102:251–318. https://doi.org/10.1007/BF03191825
4. Ho W (2008) Integrated analytic hierarchy process and its applications—A literature review. Eur J Oper Res 186:211–228. https://doi.org/10.1016/j.ejor.2007.01.004
5. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
6. Hwang C-L, Yoon K (1981) Multiple Attribute Decision Making: Methods and Applications. Springer. https://doi.org/10.1007/978-3-642-48318-9
7. Ma L, Liu H, Bai L (2017) Fire risk assessment of historic buildings based on AHP and entropy weighting. J Xi'an Univ Sci Technol 37:537–543. https://doi.org/10.13800/j.cnki.xakjdxxb.2017.0413
8. Yin P (2012) High-speed railway and tourism: Review and experience. Tourism Tribune 27:33–40
9. Hyndman RJ, Athanasopoulos G (2018) Forecasting: Principles and Practice, 2nd ed. OTexts, Available: https://otexts.com/fpp2/
10. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
11. National Bureau of Statistics of China National data and statistical releases. Available: https://data.stats.gov.cn/
12. RESSET Financial Research Database Aggregate indicator exports used under provider access terms. Available: https://www.resset.com/

Figures

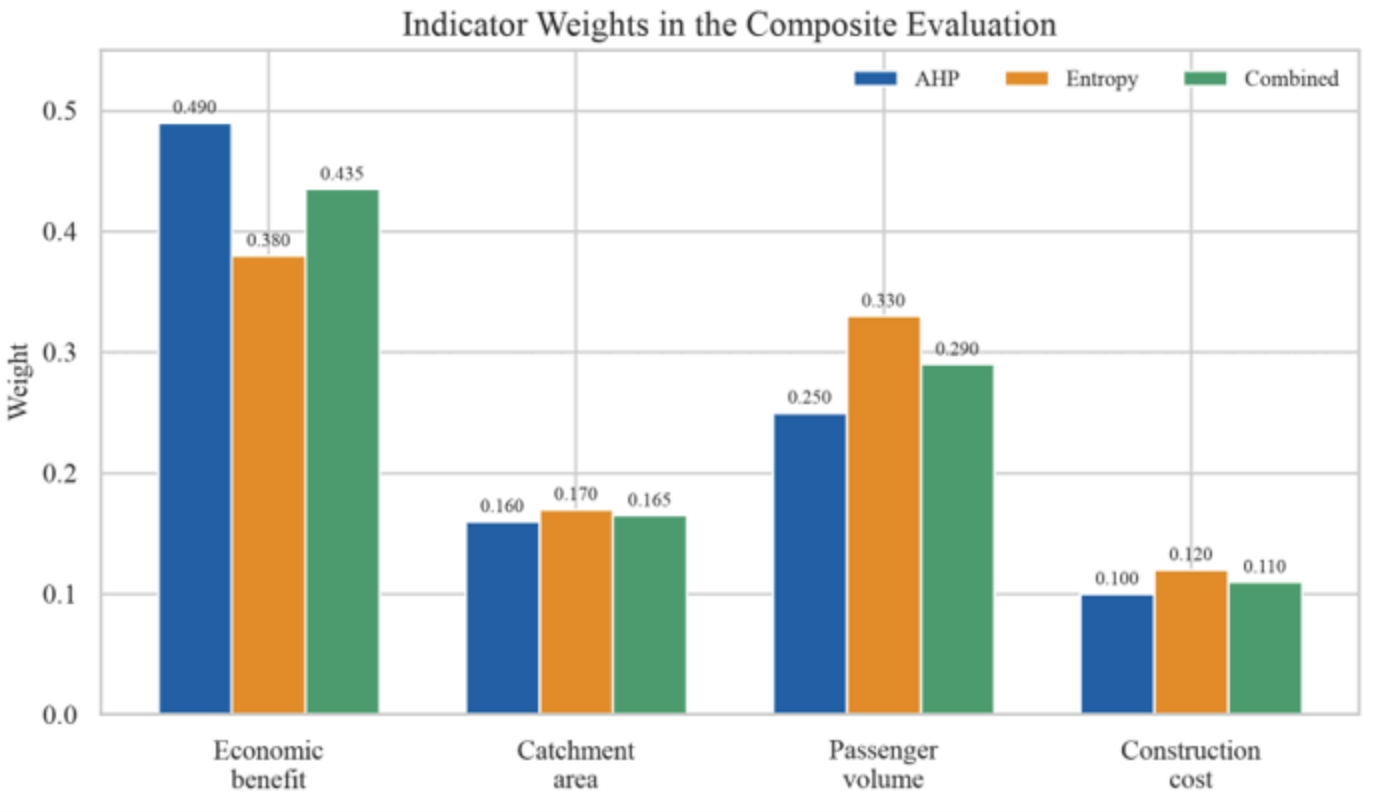


Figure 1

AHP, entropy, and combined indicator weights.

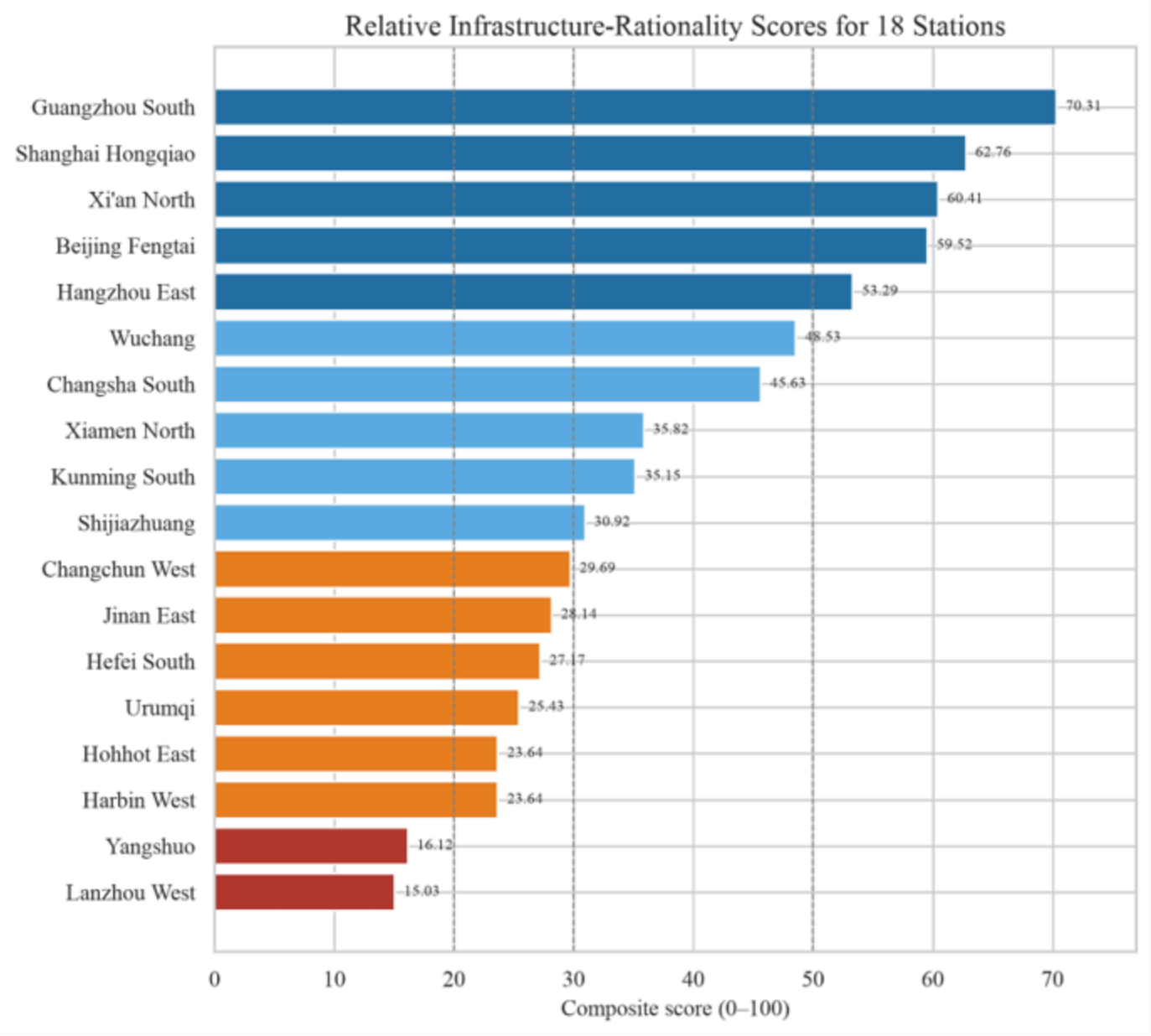


Figure 2

Relative composite scores for 18 complete-case stations.

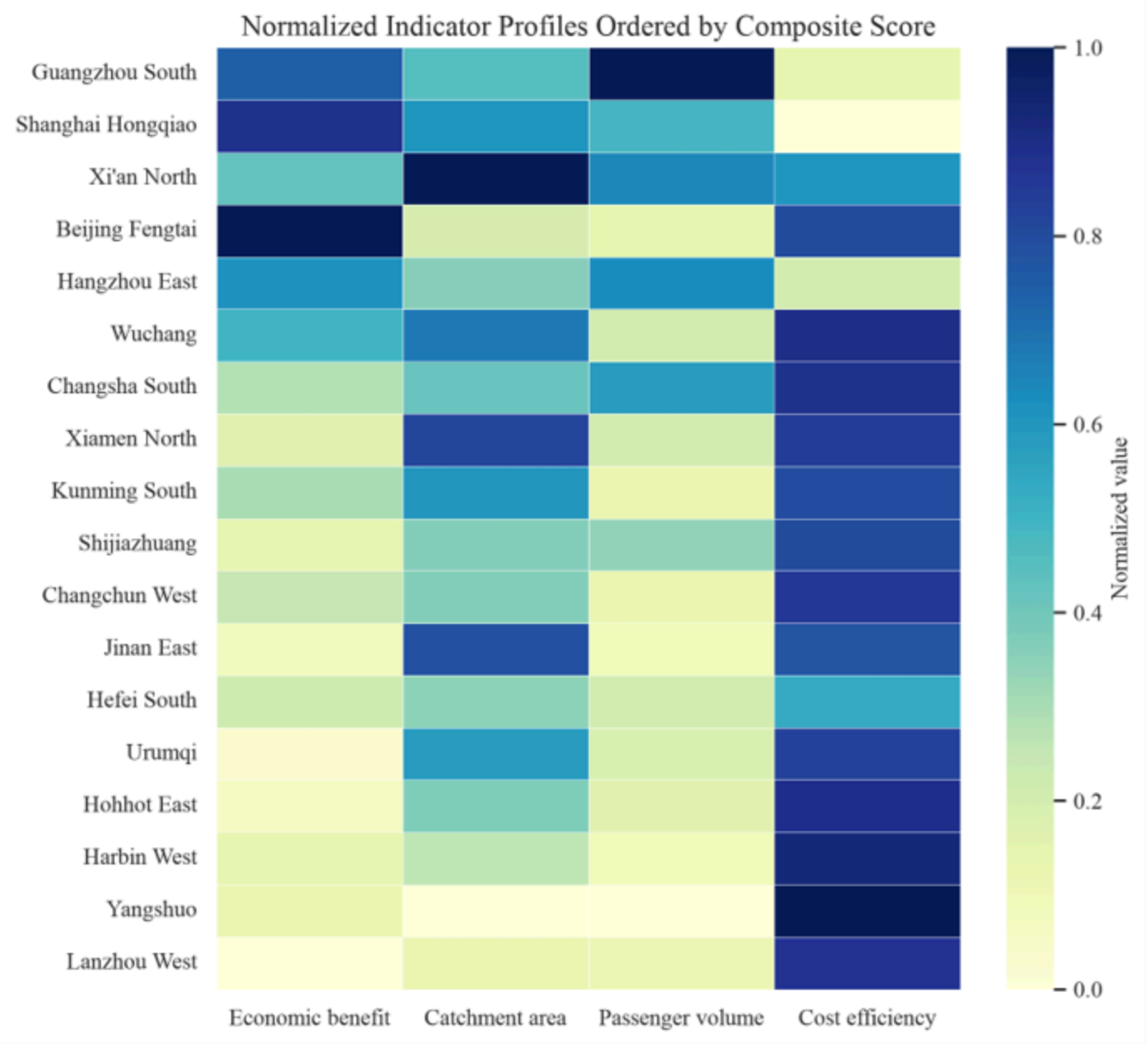


Figure 3

Normalized indicator profiles ordered by composite score.

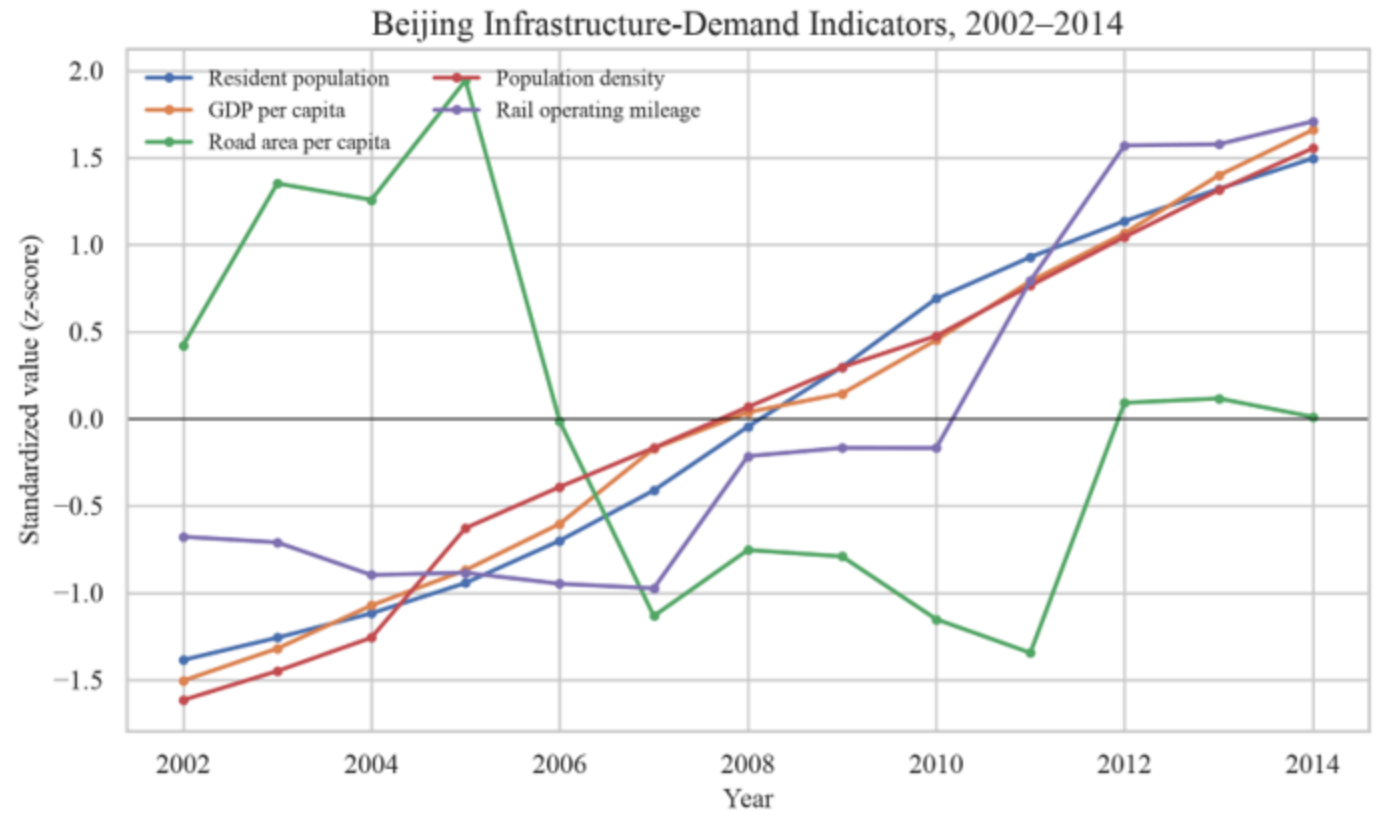


Figure 4

Standardized Beijing infrastructure-demand indicators, 2002–2014.