Environmental Research Letters

A novel global urban typology framework for sustainable mobility futures

Environmental Research Letters

OPEN ACCESS

RECEIVED
8 July 2020

ACCEPTED FOR PUBLICATION
15 July 2020

PUBLISHED
7 September 2020

Corrigendum: A novel global urban typology framework for sustainable mobility futures (2019 *Environ. Res. Lett.* 14 095006)

Jimi B Oke[1,2], Youssef M Aboutaleb[1], Arun Akkinepally[1], Carlos Lima Azevedo[1,3], Yafei Han[1], P Christopher Zegras[4], Joseph Ferreira[4] and Moshe E Ben-Akiva[1]

1 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, United States of America
2 Department of Civil and Environmental Engineering, University of Massachusetts, Amherst MA 01003, United States of America
3 Department of Management Engineering, Technical University of Denmark, Bygningstorvet, 2800 Kgs. Lyngby, Copenhagen, Denmark
4 Department of Urban Studies and Planning, Massachusetts Institute of Technology Cambridge, MA 02139, United States of America

E-mail: oke@mit.edu and jboke@umass.edu

In the first paragraph of the Introduction (page 1), the emissions due to transportation as (IEA 2017) were incorrectly reported. The corrected sentence is:

> Transportation remains a key driver for carbon dioxide emissions, accounting for nearly a quarter of the 32.3 GtC0_2-e emissions from fuel combustion globally in 2015 (IEA 2017).

In the same paragraph, estimates from the International Transport Forum report (ITF 2017) were incorrectly reported. The corrected sentence is:

> The number of motorized urban passenger-kilometers is expected to nearly double to 48.4 trillion in 2050, while the number of motor vehicles on the road is estimated to grow to 2.4 billion within the same period, from the 2015 level of 1 billion (ITF 2017).

Furthermore, in describing the MassTransit Moderate typology in the last row of table 6 (page 11), '*low* congestion' should be '*moderate* congestion.' Finally, in the last sentence of section 3.2.4 (page 11), there was a typographical error. The corrected sentence is:

> The *Hybrid Moderate* typology is largely represented in South America (e.g. Havana, Cordoba, Panama City) and Central Asia, while the *Hybrid Giant* cities are chiefly found in Eastern Europe and East Asia (e.g. Daegu, Hiroshima, Sofia).

We stress that these typographical and reporting errors were limited to the paper, and have no bearing on the modeling framework presented in the manuscript.

ORCID iD

Jimi B Oke https://orcid.org/0000-0001-6610-445X

References

IEA 2017 CO_2 emissions from fuel combustion 2017—highlights *Technical report* Int. Energy Agency Paris, France

ITF 2017 ITF Transport Outlook 2017 *Technical report* Int. Transport Forum Paris (https://doi.org/10.1787/e979b24d-en)

Environmental Research Letters

OPEN ACCESS

RECEIVED
31 December 2018

REVISED
4 April 2019

ACCEPTED FOR PUBLICATION
20 May 2019

PUBLISHED
3 September 2019

LETTER

A novel global urban typology framework for sustainable mobility futures

Jimi B Oke[1,4], Youssef M Aboutaleb[1], Arun Akkinepally[1], Carlos Lima Azevedo[1,2], Yafei Han[1], P Christopher Zegras[3], Joseph Ferreira[3] and Moshe E Ben-Akiva[1]

[1] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, United States of America
[2] Department of Management Engineering, Technical University of Denmark, Bygningstorvet, Building 116, room 121C, 2800 Kgs. Lyngby, Copenhagen, Denmark
[3] Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America
[4] Author to whom any correspondence should be addressed.

E-mail: oke@mit.edu

Keywords: urban typologies, sustainable mobility, cities, clustering, data mining, factor analysis

Abstract
Urban mobility significantly contributes to global carbon dioxide emissions. Given the rapid expansion and growth in urban areas, cities thus require innovative policies to ensure efficient and sustainable mobility. Urban typologies can serve as a vehicle for understanding dynamics of cities, which exhibit high variability in form, economic output, mobility behavior, among others. Yet, typologies relevant for sustainable urban mobility analyses are few, outdated and not large enough in scope. In this paper, we present a new typologization spanning 331 cities in 124 countries. Our sample represents 40% of the global urban population and contains the most recent data from 2008 to date. Using a factor analytic and agglomerative clustering approach, we identify 9 urban factors and 12 typologies. We discuss the implications of this new framework for researchers and planners and investigate the relationships between mobility and environmental sustainability indicators. Notably, we show an immediate application of the urban typologies to better understanding travel behavior and also describe their usage for detailed large-scale simulation in representative prototype cities for insights into sustainable future mobility policy pathways. Our data and results are publicly available for further exploration and will serve as a foundation for future analyses toward desirable urban and environmental outcomes.

1. Introduction

The rapid pace of growth in urban populations and vehicle ownership worldwide has led to an increasing demand for mobility and its associated impacts. Transportation remains a key driver for carbon dioxide emissions, accounting for nearly a quarter of the 32.3 MtCO_2-e emissions from fuel combustion globally in 2016 (IEA 2017). Urban passenger traffic further accounts for about 30% of this contribution. The number of passenger-kilometers driven is expected to more than double to 100 billion in 2050, while the number motor vehicles on the road is estimated to grow to 2.5 billion within the same period, from a current level of 1 billion. Even more critically, CO_2 emissions are projected to increase by 60% in 2050 in the absence of concrete mitigation measures (ITF 2017).

To tackle mounting environmental challenges, metropolitan and city agencies are making greater efforts to develop and share resources for understanding city dynamics. Instruments arising from these efforts can facilitate the impact assessment of future mobility solutions at the city level. To this end, various programs have been established. We note the following key examples. The International Council for Local Environmental Initiatives[5] was founded in 1990 and has influenced over 1500 urban areas worldwide. The city vitality and sustainability[6] program was

[5] https://iclei.org/.
[6] https://civitas.eu/.

established in 2002 to promote cleaner transportation in European cities. Created in 2001, the transportation planning capacity building program[7] coordinates information sharing and decision-making among planners in the US toward more efficient and sustainable mobility. Another noteworthy example is the Smart City Challenge[8] launched by the US Department of Transportation in 2015. Under this initiative, 78 cities proposed technology-driven solutions to address urban mobility challenges and reduce carbon emissions, with implementations currently underway in 5 of them (U.S. DOT 2017). The unofficial Chinese four-tier city classification system (Bland and Hernandez 2016) highlights longstanding efforts to coordinate policy for development and mobility in the world's most populous country (Li 2007). The example of China also signifies its increasingly progressive approach in developing accessible and resilient transportation systems.

Yet, the above initiatives often suffer from a lack of comprehensive and quantitative global analyses of city dynamics and their heterogeneity. This type of analysis is especially critical in today's globalized transportation market, where solutions pioneered in one city are swiftly deployed in others with mixed outcomes. Consequently, results from academic efforts in understanding functional patterns in city dynamics can be harnessed to improve outcomes for sustainable urban mobility. Over the past 80 years, researchers have approached this challenge by classifying cities chiefly in terms of geography, urban form and economy. Price (1942) proposed using factor analysis to reduce the dimensions of 15 socioeconomic and demographic variables across 93 US cities. Four key factors were identified to ultimately aid the prediction of social change. Harris (1943) was then the first to empirically classify cities, with his seminal work grouping US metropolitan areas based on primary economic *function*: manufacturing, diversification, transportation, tourism, wholesale, retail, education and mining. Bruce and Witt (1971) also applied factor analysis in their classification, producing 13 economic types of US cities from 6 factors, following earlier efforts by Hadden and Borgatta (1965).

The economic classification of cities continues to be an active topic with recent work still being put forward by researchers, enhancing the methods used and enlarging the scope of analysis. Martin *et al* (2008) proposed a 15 indicator-based socioeconomic classification of 300 cities, advancing the idea of a 'canonical city' to allow for an 'economy of effort' in policy analyses. Similarly, the 'City 600' project conducted by the McKinsey Global Institute (Dobbs *et al* 2011) attempted to track the economic power of established and emerging global cities from 2007 to 2025 by ranking them on age, GDP and household income, although no typologies are generated.

The economic classification was soon followed by urban form analysis. Huang *et al* (2007), for example, clustered 77 cities using seven spatial metrics to typologize urban form in order to provide a framework for urban development analyses. More recently, these urban form dynamics were linked to sustainability and environmental indicators. Le Néchet (2012) analyzed morphological and functional indicators of 34 European cities in order to find measures of sustainable urban form. The study further determined that energy consumption is positively correlated to wealth, automobile density, sprawl, diffusion, and polycentricity. In a recent analysis of 274 cities, notable for its scope and relevance to sustainability, Creutzig *et al* (2015) obtained 8 typologies of energy consumption, using socioeconomic and environmental indicators.

In contrast to the existing literature above, very few studies have integrated and focused on the transportation dimension in large-scale urban classification. Thomson (1977) collected and analyzed data from 30 megacities and in the process defined 5 archetypes that broadly captured the urban characteristics of the cities. We note Priester *et al* (2013) and Zegras and Gakenheimer (2005) who analyzed 41 megacities on a global scale to determine future mobility characteristics. From 59 indicators, Priester *et al* (2013) obtained 13 factors, including congestion, taxi traffic, public transit usage, parking charges, among others. Seven typologies emerged, namely: paratransit, auto, non-motorized, hybrid, traffic-saturated, transit and the singleton, Manila. However, this work is based on data collected in 1995[9] and did not include environmental or economic factors. Subsequently, McIntosh *et al* (2014) investigated how urban form influenced car ownership from 1960 through 2000. This study incorporated prior typology information but was only limited to 26 cities.

To effectively address urban efficiency and environmental concerns, a mobility-oriented global urban typologization based on recent relevant data is required. As noted, the majority of existing comprehensive urban typology work has been limited to a few regions: Europe, China and the US. Those with a global scope are either now outdated or do not account for transportation-related variables. Given the significant contribution of mobility to CO_2 emissions and consequently climate change, effective pathways to sustainability must include sufficiently detailed transportation variables. Most critically, there is a dearth of publicly available data to inform decisions by planners and policymakers. The new urban typologies we have discovered in our effort target the aforementioned gaps. Our analyses span the most recent data from 331 cities worldwide, which is the largest in scope

[7] https://planning.dot.gov/focus_metropolitan.asp
[8] https://transportation.gov/smartcity

[9] The UITP Millenium Cities Database for Sustainable Transport (Kenworthy and Laube).

Table 1. Summary of data sources, indicators, years and number of cities.

Source(s)	Indicators (units)	Years	No. cities
Demographia (2017)	Population, land area (km^2), population density (per km^2)	2016	331
Global BRTData (2017)	Fleet size (per 100 K), fare, stations (per 100 K), system length (km), annual ridership	2010–17	301–331
Global Urban Indicators Database (2015)	GDP (USD), poverty rate (%), life expectancy (years)	2013	93–331
Global Petrol Prices (2017)	Gasoline price (USD)	2017	331
Innovation Cities Index (2017)	Innovation score	2015	238
Internet World Stats (2017)	Internet penetration (%), digital access (%)	2017	323–329
Numbeo (2017)	Urban indices: cost of living, rent, groceries, purchasing power, affordability, safety, pollution, traffic (time), inefficiency, emissions	2016	126–223
OpenStreetMap (2017)	Circuity average, degree average, intersections, intersection density (per km^2) street length (km), street length average (km), street length density (per km) self-loop proportion, highway proportion	2017	243–259
Pew Research Center (2016)	Smartphone penetration (%)	2017	218
TomTom (2016)	Congestion level: overall, morning peak, evening peak (%)	2016	146–154
UN-Habitat (2015)	Population, Gini coefficient, CO_2 emissions (metric tons per capita), unemployment, urbanization level	2011–2014	129–331
World Health Organization (2013)	Road traffic deaths	2013	313

in the academic literature to the best of our knowledge. We incorporate economic, demographic, urban form, mobility and environmental indicators. Our compiled data and results are freely available and applicable to (a) further understanding city dynamics and behavior, and (b) enabling the coherent impact analysis of future scenarios towards sustainable urban mobility.

2. Methods

For the purposes of this urban typologization, we defined a 'city' as an urban agglomeration with a population of at least 750 000. Over 700 cities met this criterion. However, owing to the ready availability of open data sources, we further reduced our city sample to 331, ensuring global and national representativity. This final sample spans 124 countries and represents 40% of the global urban population. We collected urban data from a variety of open sources, with the aim of representing the most recent snapshot, thus limiting the variables to no earlier than 2008. Following data collection, we conducted an exploratory factor analyses to obtain latent urban attributes and consequently reduce dimensionality for further differentiation. Finally, we applied clustering methods, using the identified attributes, to obtain the typologies. We validated our results by examining typology characteristics across the factors and key variables.

2.1. Data

The data collected for this study consists of 64 indicators across seven urban dimensions: mobility, economy, environment, social development, urban form and geography. We compiled these data from public and open sources and have made them publicly available (along with necessary metadata) at http://its.mit.edu/typologies. Key sources and variables are summarized in table 1. To obtain the most recent and accurate mode share data, metropolitan, regional and national documents were accessed.

We validated and cleaned the dataset by inspection. In cases where appropriate and where city-level data were unavailable, urban average values at the country level were used. Network data was obtained from OpenStreetMap. Given the inconsistencies in the boundaries of metropolitan regions, we computed the network statistics (Boeing 2017) based on the urban core of each city, which we found to be consistent across all the available cities. To prepare the data for factor analyses, each variable was scaled to a maximum value of 10.

2.2. Factor analysis

We conducted an exploratory factor analysis (EFA) to unearth the latent structure of the dataset and also to reduce dimensionality for the clustering which followed. Indicators related to metro, BRT and bikeshare were treated as left-censored variables truncated at 0, given that a sizable fraction of the cities do not have these services available. Thus, we used the generalized Tobit factor analysis approach, defining the vector of J

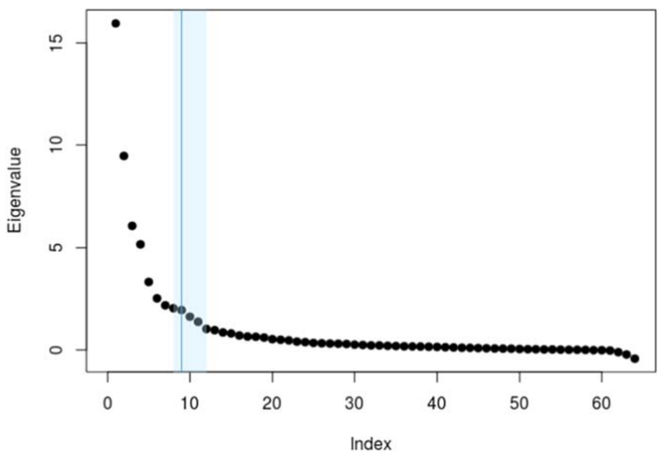

Figure 1. Eigenvalues of the sample correlation matrix. The blue vertical line at the 9th index indicates the number of factors that was ultimately extracted. The shaded rectangle indicates the knee region which serves as a guide to the number of factors that would best represent the sample data.

latent response variables $\boldsymbol{y}^*$, where the actual measurement is given by y_j in each case. The factor analysis model is thus:

$$\boldsymbol{y}^* = \boldsymbol{\nu} + \boldsymbol{\Lambda\eta} + \varepsilon, \tag{1}$$

$$y_j = \begin{cases} y_j^*: & y_j^* > 0 \\ 0: & y_j^* \leqslant 0 \end{cases}, \tag{2}$$

where $\boldsymbol{\nu}$ is the vector of variable means, $\boldsymbol{\Lambda}$ a $J \times P$ matrix of factor loadings; P is the number of factors, $\boldsymbol{\eta}$ the $P \times 1$ vector of variable scores on each factor, and ε the $J \times 1$ vector of independent error terms The covariance structure of $\boldsymbol{y}^*$ is given by

$$\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{y}^*) = \boldsymbol{\Lambda}'\boldsymbol{\Psi\Lambda} + \boldsymbol{\Theta}, \tag{3}$$

where $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ are the factor and error covariance matrices, respectively. Further details on the formulation of this factor analytical approach can be found in Muthen (1989) and Kamakura and Wedel (2001).

A maximum likelihood approach was used to estimate the model. Thus, standard errors and a test statistic were computed as robust to non-normality and non-independence of observations. To handle missing data, full-information maximum likelihood was incorporated. In the estimation procedure, the conditional distributions of variables with missingness were therefore assumed to be multivariate normal on variables with complete observations across the sample, ensuring that all observations were utilized in model estimation (Hirose *et al* 2016, Cham *et al* 2017). The rate of missingness in the dataset is 15%, and we assumed that observations were missing at random.

We note that the goal of EFA is to produce an interpretable parsimonious representation with similar correlation structure to that obtained from the original sample. Therefore, as an initial estimate of the appropriate number of factors required, we consider the eigenvalues of the sample correlation matrix. This scree plot is shown in figure 1. From this, we see that extracting any number between 8 and 12 factors will sufficiently capture the variance in the dataset.

2.2.1. Model fitness

To formally judge model fitness, we used likelihood-based approaches. We first employed the likelihood-ratio test (LRT) to successively test m-factor models against $(m + 1)$-factor models. The LRT has been shown to have limitations in that it could lead to overfactoring. However, this property is not undesirable, as the greater the number of factors, the closer the estimated structure will be to the true structure (Hayashi *et al* 2007). These test statistics are shown in table 2. χ^2_{diff} is the LRT statistic (Satorra and Bentler 2001) used to test the null hypothesis (H_0) that the data structure has at most m factors against the alternative of at least $m + 1$ factor. In all successive models the test statistic is significant enough to reject H_0. Clearly, we cannot rely on the LRT alone in this case for the most parsimonious choice. However the relative sizes of χ^2_{diff} indicate that within the current range, the larger models will not drastically improve the fit.

As has been long established (Jöreskog 1969), statistical testing alone is inadequate to make a final decision. Given that the AIC and BIC both decreased as the number of factors increased to 11, we also had to consider interpretability and simplicity to guide the model selection between 8 and 11 factors. Thus, we computed the loading simplicity index (LSI) (Lorenzo-Seva 2003), which measures the simplicity of a factor solution on a scale of 0 (most complex) to 1 (simplest). These values are shown as well in table 2, providing further support for the 9-factor model. LSI* denotes the case in which only the statistically significant loadings are used. Finally, we visually inspected each solution and chose the 9-factor model as it gave the most parsimonious and interpretable results.

Table 2. Fit statistics for the Tobit factor models. The 9-factor model was deemed the best fit, taking into account simplicity and interpretability.

No. of factors	Log-likelihood	AIC	BIC	χ^2_{diff}	df_{diff}	LSI	LSI*
3	−24 877.3	50 388.5	51 593.8	—	—	0.380	—
4	−24 029.4	48 814.7	50 251.9	1035.9	61	0.401	0.655
5	−23 447.4	47 770.9	49 436.2	975.9	60	0.397	0.616
6	−23 005.2	47 004.4	48 894.0	743.8	59	0.416	0.693
7	−22 512.6	46 135.2	48 245.4	819.6	58	0.481	0.671
8	−22 115.5	45 455.0	47 781.9	500.6	57	0.504	0.727
9	**−21 965.3**	**45 266.7**	**47 806.5**	**308.2**	**56**	**0.531**	**0.770**
10	−21 672.8	44 791.5	47 540.5	699.4	55	0.523	0.770
11	−21 492.2	44 538.4	47 492.6	321.7	54	0.534	0.830
12	−21 390.6	44 441.1	47 597.0	206.8	53	0.536	0.784

The factor loadings $\boldsymbol{\Lambda}$ were obliquely rotated using the Geomin method (Yates 1988) for optimal interpretation. We based our choice of the oblique rotation on the hypothesis that the latent factors are correlated. Further, Geomin has been shown to provide stable loadings and better results when the original loading pattern is unknown, as was the case here (Celimli Aksoy 2017). The objective of the rotation is to maximize interpretability by minimizing the criterion $\boldsymbol{Q}(\boldsymbol{\Lambda}) = \sum_J (\prod_P (\lambda_{jp}^2 + \varepsilon))^{\frac{1}{P}}$ (Hattori *et al* 2017) using an iterative approach that begins with a random starting point. λ_{jp} are the elements of the factor loading matrix, while ε is a small number added to prevent indeterminacy when $\lambda_{jp} = 0$. We used the recommended 100 random starts with a maximum of 10 000 iterations to find the best local minimum. In the objective function, the rotated loading matrix $\boldsymbol{\Lambda} = \boldsymbol{AT}$, where $\boldsymbol{A}$ is the unrotated loading matrix. Thus, the solution finds the optimal rotation matrix $\boldsymbol{T}$. The rotated factor loadings and standard errors are shown in table 3. Factor correlations are shown in table 4 [10].

Finally, we computed the factor scores, i.e. the reduced dataset, using the components method. Specifically, we found the weight matrix $\boldsymbol{W} = \boldsymbol{R}^{-1}\boldsymbol{\Lambda}^*$, where $\boldsymbol{R}$ is the sample correlation matrix. The reduced dataset was then given by $\tilde{\boldsymbol{Y}}_{331\times 9} = \boldsymbol{Y}_{331\times 64}\boldsymbol{W}_{64\times 9}$.

2.3. Hierarchical agglomerative clustering

Hierarchical clustering has been shown to be effective at pattern recognition, particularly because it is unsupervised and thus requires no predefined assumptions on the nature of the data (Jain *et al* 1999, Jain 2010). Given a separation metric between elements in a multidimensional set, hierarchical clustering proceeds by iteratively agglomerating groups based on various criteria (e.g. Ward's method, method of averages (UPGMA), single linkage, among others). Using the 331 × 9 factor score matrix obtained from the factor analytical procedure, we computed a 331 × 331 dissimilarity matrix based on the Manhattan metric. Upon comparing the results of several agglomerative approaches, we finally selected Ward's method (Murtagh and Legendre 2014) to cluster the cities based on the factor score dissimilarities. Ward's method has historically been shown to consistently perform best in a variety of applications (Kuiper and Fisher 1975, Blashfield 1976, Hands and Everitt 1987, Milligan and Cooper 1988, Ferreira and Hitchcock 2009). In this case, it produces the most balanced clusters with assignments adjudged to be the most valid.

In addition to selecting the clustering method, the other important question was the number of relevant clusters in the dataset. Various measures and heuristics have been proposed to elicit the cluster number that produces the optimal separation of members into groups, but the objective of interpretability ultimately dictate the final choice. We compared the results of 30 clustering validity indices using the `NbClust` package in R (Charrad *et al* 2014). With the minimum number of clusters set to 4, nine of the indices (including Hartigan, Tau and Ratkowsky) indicated that 5 was the optimal number of clusters. As we increased the minimum cluster number constraint, successively optimal cuts were found at 6, 8, 9, 11 and 13 clusters, respectively, as can be seen in the dendrogram (figure 2).

We observed that a 6-cluster result produced the higher-level typologies as described in section 3.2, namely: *Auto*, *BusTransit*, *Congested*, *Hybrid*, *MetroBike*, and *MassTransit*. The 13-cluster result splits the *MassTransit Heavyweight* into two further groups which are distinguished by population size. Moving further up the tree to 12 clusters produced the pairwise members of all the aforementioned higher-level groups. Given the interpretative benefit of the 12 cluster solution, however, we chose this as the final result. Our decision was supported by a precedent for not always following the majority rule in choosing the optimal cluster number (Charrad *et al* 2014). After the first clustering solution was obtained, we iterated by imputing the cluster means into the dataset at the factor score computation stage. This iterative process was repeated until the cluster results were stable in order to obtain the final typology assignments.

[10] The exploratory factor analysis was conducted in Mplus v.8 using the MLR estimator (Muthén and Muthén 2017).

Table 3. Loadings of each variable onto the 9 factors.

Variable	Loading (Standard Error)																	
	Metro propensity		BRT propensity		Bikeshare propens.		Development		Sustainability		Population		Congestion		Sprawl		Network density	
Metro length	0.862[a]	(0.026)	−0.046[a]	(0.017)	0.062[a]	(0.026)	−0.042	(0.031)	0.009	(0.013)	0.081[a]	(0.019)	0.011	(0.022)	0.059[a]	(0.018)	0.077[a]	(0.036)
Metro length density	0.870[a]	(0.024)	0.021	(0.023)	−0.014	(0.024)	0.061[a]	(0.029)	0.037	(0.019)	−0.007	(0.021)	0.043	(0.030)	0.021	(0.037)	−0.105[a]	(0.034)
Metro stations per 100K	0.825[a]	(0.029)	0.069[a]	(0.019)	0.159[a]	(0.040)	0.001	(0.021)	−0.014	(0.016)	−0.056[a]	(0.020)	−0.073	(0.041)	−0.074[a]	(0.019)	−0.033	(0.027)
Metro ridership p.c.	0.861[a]	(0.023)	0.032	(0.021)	0.035	(0.033)	0.044	(0.024)	0.023	(0.013)	0.024	(0.017)	0.045	(0.029)	−0.016	(0.027)	−0.019	(0.027)
Metro age	0.785[a]	(0.029)	0.096[a]	(0.024)	0.080[a]	(0.040)	0.054	(0.031)	−0.088[a]	(0.022)	0.021	(0.023)	−0.01	(0.026)	0.003	(0.025)	0.091[a]	(0.039)
BRT length	0.039	(0.027)	0.864[a]	(0.023)	0.014	(0.039)	0.064	(0.036)	−0.058[a]	(0.025)	0.142[a]	(0.026)	0.009	(0.029)	0.001	(0.020)	−0.009	(0.021)
BRT system length density	−0.023	(0.023)	0.903[a]	(0.027)	0.112[a]	(0.046)	0.011	(0.028)	0.016	(0.021)	0.065[a]	(0.027)	0.018	(0.028)	−0.014	(0.021)	−0.071[a]	(0.023)
BRT stations per 100K	−0.054[a]	(0.024)	0.860[a]	(0.025)	0.086	(0.050)	0.119[a]	(0.038)	0.027	(0.023)	0.045	(0.026)	−0.022	(0.026)	−0.017	(0.020)	−0.038	(0.021)
BRT fleet per 100K	0.079[a]	(0.035)	0.926[a]	(0.024)	0.019	(0.031)	0.026	(0.036)	−0.032	(0.017)	0.023	(0.021)	0.001	(0.022)	−0.02	(0.019)	−0.022	(0.026)
BRT annual ridership p.c.	0.044	(0.028)	0.983[a]	(0.026)	0.001	(0.030)	−0.025	(0.030)	0.003	(0.022)	−0.006	(0.023)	0.048	(0.028)	0.032	(0.025)	0.02	(0.019)
BRT age	0.076[a]	(0.038)	0.883[a]	(0.029)	−0.01	(0.040)	0.082[a]	(0.038)	−0.01	(0.030)	0.036	(0.025)	0.036	(0.030)	−0.021	(0.027)	0.062	(0.032)
Bikeshare stations	0.054	(0.029)	0.085[a]	(0.035)	0.819[a]	(0.060)	0.073	(0.062)	0.056	(0.049)	0.041	(0.037)	0.025	(0.034)	0.022	(0.032)	0.081[a]	(0.040)
Bikeshare stations per 100K	0.024	(0.030)	0.092[a]	(0.038)	0.815[a]	(0.064)	0.105	(0.068)	0.057	(0.048)	−0.011	(0.031)	0.003	(0.031)	−0.022	(0.033)	0.043	(0.036)
Bikeshare number of bikes	0.025	(0.025)	0.054	(0.032)	0.879[a]	(0.051)	0.035	(0.057)	0.074	(0.054)	0.062	(0.033)	0.022	(0.042)	0.054	(0.032)	0.066	(0.037)
Bikeshare bicycles per 100K	0.008	(0.024)	0.055	(0.032)	0.896[a]	(0.055)	0.027	(0.061)	0.088	(0.054)	0.009	(0.036)	0.011	(0.038)	0.045	(0.030)	0.05	(0.036)
Bikeshare age	0.074	(0.039)	0.03	(0.044)	0.686[a]	(0.072)	0.257[a]	(0.077)	−0.005	(0.038)	0.067	(0.039)	0.071	(0.044)	−0.063	(0.043)	0.031	(0.042)
Car modeshare	−0.203[a]	(0.054)	0.017	(0.027)	0.333[a]	(0.121)	0.437[a]	(0.123)	−0.269[a]	(0.061)	−0.044	(0.048)	−0.337[a]	(0.165)	0.208	(0.123)	−0.028	(0.039)
Public transit modeshare	0.203[a]	(0.075)	0.001	(0.097)	−0.216	(0.280)	−0.264	(0.172)	0.073	(0.108)	−0.06	(0.088)	0.360[a]	(0.140)	−0.149	(0.113)	−0.001	(0.064)
Bicycle modeshare	−0.018	(0.055)	0.036	(0.084)	−0.115	(0.177)	0.009	(0.105)	0.674[a]	(0.076)	0.216	(0.150)	0.031	(0.085)	0.024	(0.085)	0.026	(0.054)
Walking modeshare	0.118	(0.078)	0.024	(0.092)	−0.275	(0.229)	−0.408[a]	(0.158)	0.141	(0.094)	0.013	(0.069)	0.2	(0.169)	−0.212[a]	(0.099)	0.072	(0.065)
Gasoline pump price	0.065	(0.065)	−0.018	(0.050)	−0.007	(0.063)	0.348[a]	(0.150)	0.186[a]	(0.079)	−0.017	(0.039)	0.167	(0.199)	−0.549[a]	(0.089)	−0.044	(0.060)
Road deaths rate	−0.01	(0.034)	0.066	(0.044)	−0.205[a]	(0.090)	−0.704[a]	(0.090)	−0.028	(0.054)	−0.091	(0.058)	0.009	(0.068)	0.375[a]	(0.043)	0.03	(0.043)
Congestion	−0.03	(0.026)	0.022	(0.040)	−0.002	(0.049)	−0.101	(0.077)	−0.024	(0.042)	0.112	(0.078)	0.870[a]	(0.041)	−0.027	(0.142)	−0.038	(0.080)
Congestion AM peak	0.005	(0.035)	−0.037	(0.044)	0.028	(0.066)	−0.065	(0.089)	−0.029	(0.046)	0.038	(0.060)	0.846[a]	(0.069)	−0.096	(0.154)	−0.086	(0.089)
Congestion PM peak	−0.005	(0.031)	0.029	(0.047)	0.07	(0.072)	−0.012	(0.059)	−0.053	(0.067)	−0.024	(0.071)	0.935[a]	(0.044)	0.11	(0.147)	−0.121	(0.111)
Traffic index	0.021	(0.036)	0.009	(0.035)	−0.112	(0.232)	−0.025	(0.057)	−0.768[a]	(0.088)	0.169	(0.177)	0.294[a]	(0.137)	0.08	(0.048)	0.01	(0.023)
Travel time index	0.098[a]	(0.038)	0.038	(0.043)	−0.137	(0.207)	0.026	(0.044)	−0.694[a]	(0.081)	0.146	(0.172)	0.396[a]	(0.117)	0.007	(0.038)	−0.005	(0.029)
Inefficiency index	0.027	(0.046)	−0.02	(0.052)	−0.134	(0.267)	0.11	(0.080)	−0.774[a]	(0.094)	0.14	(0.177)	0.225	(0.148)	0.081	(0.061)	−0.039	(0.032)
Population	0.088[a]	(0.038)	0.032	(0.038)	0.093	(0.050)	0.148	(0.079)	−0.029	(0.048)	0.605[a]	(0.080)	0.238[a]	(0.089)	0.016	(0.038)	0.215[a]	(0.070)
Land area	0.112[a]	(0.050)	−0.083	(0.047)	−0.004	(0.050)	0.518[a]	(0.103)	−0.03	(0.041)	0.292[a]	(0.073)	0.129	(0.146)	0.259[a]	(0.056)	0.280[a]	(0.082)
Population density	0.003	(0.045)	−0.015	(0.048)	−0.046	(0.072)	−0.392[a]	(0.108)	−0.075	(0.053)	0.395[a]	(0.092)	0.075	(0.085)	−0.240[a]	(0.060)	−0.151[a]	(0.059)
Population Δ 1990-00	0.042	(0.028)	0.067[a]	(0.031)	0.041	(0.062)	0.033	(0.045)	0.141	(0.081)	0.892[a]	(0.032)	−0.029	(0.044)	0.025	(0.029)	0.059	(0.042)
Population Δ 2000-10	0.014	(0.024)	0.063[a]	(0.032)	0.032	(0.036)	0.033	(0.034)	0.099	(0.076)	0.949[a]	(0.024)	−0.019	(0.041)	0.014	(0.024)	0.071	(0.040)
Population Δ 2010-20	0.022	(0.020)	0.048[a]	(0.024)	0.037	(0.024)	−0.021	(0.023)	0.03	(0.074)	0.975[a]	(0.016)	−0.003	(0.027)	−0.032	(0.030)	0.028	(0.031)

Table 3. (Continued.)

Variable	Loading (Standard Error)																	
	Metro propensity		BRT propensity		Bikeshare propens.		Development		Sustainability		Population		Congestion		Sprawl		Network density	
Population Δ 2020-25	−0.005	(0.023)	0.007	(0.030)	0.032	(0.035)	−0.076	(0.043)	−0.14	(0.075)	0.952[a]	(0.042)	0.011	(0.034)	−0.103[a]	(0.039)	−0.041	(0.033)
Urbanization rate 2015	0.055	(0.049)	0.265[a]	(0.053)	0.02	(0.074)	0.483[a]	(0.077)	0.027	(0.050)	−0.341[a]	(0.061)	−0.083	(0.092)	0.111	(0.067)	0.141[a]	(0.053)
Urbanization rate Δ 2015-25	−0.025	(0.039)	0.017	(0.047)	−0.105	(0.108)	−0.333[a]	(0.078)	0.483[a]	(0.069)	0.459[a]	(0.099)	0.034	(0.059)	0.134	(0.087)	−0.043	(0.050)
GDP per capita	0.067	(0.035)	−0.08	(0.041)	0.005	(0.053)	0.931[a]	(0.048)	−0.041	(0.029)	−0.017	(0.030)	−0.092	(0.071)	−0.01	(0.044)	−0.04	(0.034)
Unemployment rate	−0.043	(0.063)	−0.151	(0.091)	−0.063	(0.085)	−0.322[a]	(0.104)	−0.399[a]	(0.089)	0.054	(0.076)	−0.087	(0.095)	−0.299[a]	(0.069)	0.015	(0.044)
Cost of living index	−0.05	(0.026)	0.047[a]	(0.021)	−0.698[a]	(0.129)	1.249[a]	(0.065)	−0.009	(0.023)	−0.026	(0.020)	0.043	(0.025)	−0.075[a]	(0.025)	0.046[a]	(0.020)
Rent index	0.008	(0.028)	−0.079	(0.056)	−0.734[a]	(0.158)	1.239[a]	(0.069)	0.006	(0.038)	0.097[a]	(0.044)	0.169[a]	(0.067)	0.021	(0.040)	0.041	(0.041)
Grocery index	−0.067	(0.035)	0.039	(0.029)	−0.667[a]	(0.127)	1.223[a]	(0.070)	0.02	(0.033)	0.019	(0.025)	0.027	(0.056)	0.023	(0.027)	0.035	(0.039)
Restaurant price index	−0.018	(0.033)	0.054[a]	(0.027)	−0.487[a]	(0.110)	1.143[a]	(0.066)	−0.078	(0.043)	−0.073[a]	(0.030)	−0.013	(0.035)	−0.128[a]	(0.036)	0.018	(0.031)
Local purchasing power index	−0.033	(0.047)	0.049	(0.050)	0.17	(0.126)	0.613[a]	(0.106)	−0.037	(0.054)	0.134[a]	(0.044)	−0.302[a]	(0.094)	0.026	(0.074)	−0.059	(0.042)
Gini coefficient	−0.018	(0.069)	0.337[a]	(0.071)	−0.211	(0.125)	−0.033	(0.092)	−0.360[a]	(0.072)	−0.086	(0.100)	−0.111	(0.090)	0.119	(0.085)	0.073	(0.066)
Poverty rate	−0.053	(0.101)	0.210[a]	(0.102)	−0.262	(0.261)	−0.216	(0.196)	−0.303[a]	(0.104)	−0.007	(0.071)	−0.164	(0.098)	0.063	(0.092)	−0.041	(0.075)
Life expectancy	0.044	(0.031)	0.230[a]	(0.051)	0.169	(0.099)	0.468[a]	(0.069)	0.286[a]	(0.055)	−0.180[a]	(0.078)	−0.057	(0.054)	0.013	(0.035)	0.061	(0.032)
Safety index	0.05	(0.069)	−0.198[a]	(0.091)	−0.001	(0.122)	0.301[a]	(0.113)	0.472[a]	(0.074)	0.017	(0.052)	0.014	(0.066)	−0.127	(0.098)	−0.15	(0.087)
Internet penetration	0.079[a]	(0.034)	0.187[a]	(0.044)	0.036	(0.062)	0.661[a]	(0.058)	0.102[a]	(0.043)	−0.195[a]	(0.048)	−0.068	(0.057)	−0.008	(0.038)	0.001	(0.035)
Digital penetration	0.065[a]	(0.023)	0.116[a]	(0.028)	0.141[a]	(0.054)	0.710[a]	(0.055)	0.026	(0.029)	−0.175[a]	(0.031)	−0.135[a]	(0.053)	0.009	(0.047)	−0.01	(0.023)
Innovation index	0.154[a]	(0.033)	0.072[a]	(0.034)	0.211[a]	(0.050)	0.661[a]	(0.056)	0.02	(0.027)	0.077[a]	(0.035)	0.076	(0.054)	0.002	(0.030)	0.042	(0.033)
Smartphone penetration	0.044	(0.042)	0.073	(0.045)	0.041	(0.066)	0.655[a]	(0.076)	0.239[a]	(0.051)	−0.006	(0.052)	−0.244[a]	(0.080)	0.120[a]	(0.055)	−0.053	(0.043)
CO_2 emissions p.c.	0.062	(0.049)	−0.069	(0.041)	0.069	(0.068)	0.561[a]	(0.086)	0.025	(0.054)	0.023	(0.033)	−0.269[a]	(0.122)	0.231[a]	(0.077)	−0.045	(0.045)
Pollution index	−0.032	(0.053)	−0.168[a]	(0.064)	0.052	(0.096)	−0.636[a]	(0.074)	0.012	(0.059)	0.242[a]	(0.047)	0.188[a]	(0.063)	0.098	(0.054)	0.109	(0.056)
Street length total	−0.082	(0.095)	0.075	(0.097)	0.026	(0.096)	0.166	(0.106)	0.111	(0.069)	0.08	(0.108)	0.16	(0.132)	0.239[a]	(0.060)	−0.055	(0.072)
Street length density	−0.04	(0.054)	0.091	(0.083)	0.245[a]	(0.098)	0.133	(0.098)	−0.239[a]	(0.069)	−0.01	(0.042)	0.022	(0.051)	−0.109	(0.120)	0.531[a]	(0.067)
Street length average	0.011	(0.037)	−0.044	(0.048)	0.001	(0.057)	−0.02	(0.048)	0.610[a]	(0.081)	0.166	(0.111)	−0.007	(0.045)	0.290[a]	(0.096)	−0.245[a]	(0.072)
Intersection count	−0.066	(0.094)	0.102	(0.091)	0.05	(0.111)	0.15	(0.117)	−0.06	(0.075)	0.089	(0.105)	0.191	(0.146)	0.157[a]	(0.065)	0.073	(0.085)
Intersection density	−0.189[a]	(0.090)	0.079	(0.091)	0.089	(0.130)	0.018	(0.076)	0.004	(0.059)	−0.111	(0.067)	−0.003	(0.086)	−0.226[a]	(0.091)	0.360[a]	(0.097)
Degree average	−0.07	(0.072)	−0.126	(0.072)	−0.323[a]	(0.117)	−0.084	(0.097)	−0.123	(0.088)	0.064	(0.076)	−0.138	(0.153)	0.013	(0.061)	0.261[a]	(0.111)
Streets per node	0.039	(0.050)	0.002	(0.048)	−0.126	(0.125)	0.03	(0.055)	0.267[a]	(0.083)	0.234	(0.121)	−0.155	(0.179)	0.062	(0.056)	0.512[a]	(0.070)
Circuity	−0.015	(0.054)	0.167[a]	(0.064)	0.034	(0.075)	0.1	(0.082)	0.069	(0.094)	−0.234[a]	(0.074)	0.139	(0.123)	0.052	(0.074)	−0.528[a]	(0.063)
Self-loop proportion	−0.097	(0.071)	0.093	(0.050)	0.007	(0.048)	0.447[a]	(0.079)	−0.132	(0.072)	0.004	(0.054)	−0.197	(0.114)	−0.047	(0.057)	−0.287[a]	(0.053)
Highway proportion	0.016	(0.045)	0.028	(0.049)	0.018	(0.062)	0.454[a]	(0.071)	0.425[a]	(0.088)	0.207[a]	(0.089)	−0.077	(0.076)	0.217[a]	(0.069)	−0.158[a]	(0.067)

[a] Indicates significance at the 5% level. Standard errors are parenthesized.

Table 4. Factor correlations with standard errors parenthesized.

Factor	Metro propensity	BRT propensity	Bikeshare propensity	Development	Sustainability	Population	Congestion	Sprawl	Network density
Metro propensity	1	0.330[a]	0.496[a]	0.360[a]	0.287[a]	0.269[a]	0.271[a]	−0.056	0.158
	(0.000)	(0.038)	(0.039)	(0.037)	(0.089)	(0.045)	(0.062)	(0.108)	(0.108)
BRT propensity	0.330[a]	1	0.536[a]	0.647[a]	0.013	0.134[a]	−0.043	0.114	0.067
	(0.038)	(0.000)	(0.046)	(0.022)	(0.093)	(0.034)	(0.105)	(0.115)	(0.106)
Bikeshare propensity	0.496[a]	0.536[a]	1	0.665[a]	0.405[a]	0.105	−0.135	0.154	−0.02
	(0.039)	(0.046)	(0.000)	(0.042)	(0.100)	(0.055)	(0.075)	(0.104)	(0.101)
Development	0.360[a]	0.647[a]	0.665[a]	1	0.073	0.011	−0.202[a]	0.121	0.153
	(0.037)	(0.022)	(0.042)	(0.000)	(0.087)	(0.036)	(0.069)	(0.179)	(0.104)
Sustainability	0.287[a]	0.013	0.405[a]	0.073	1	−0.048	−0.055	−0.024	−0.123
	(0.089)	(0.093)	(0.100)	(0.087)	(0.000)	(0.106)	(0.086)	(0.062)	(0.094)
Population	0.269[a]	0.134[a]	0.105	0.011	−0.048	1	0.438[a]	0.172	0.104
	(0.045)	(0.034)	(0.055)	(0.036)	(0.106)	(0.000)	(0.066)	(0.134)	(0.107)
Congestion	0.271[a]	−0.043	−0.135	−0.202[a]	−0.055	0.438[a]	1	−0.168	−0.028
	(0.062)	(0.105)	(0.075)	(0.069)	(0.086)	(0.066)	(0.000)	(0.138)	(0.114)
Sprawl	−0.056	0.114	0.154	0.121	−0.024	0.172	−0.168	1	0.024
	(0.108)	(0.115)	(0.104)	(0.179)	(0.062)	(0.134)	(0.138)	(0.000)	(0.091)
Network density	0.158	0.067	−0.02	0.153	−0.123	0.104	−0.028	0.024	1
	(0.108)	(0.106)	(0.101)	(0.104)	(0.094)	(0.107)	(0.114)	(0.091)	(0.000)

[a] Indicates significance at the 5% level.

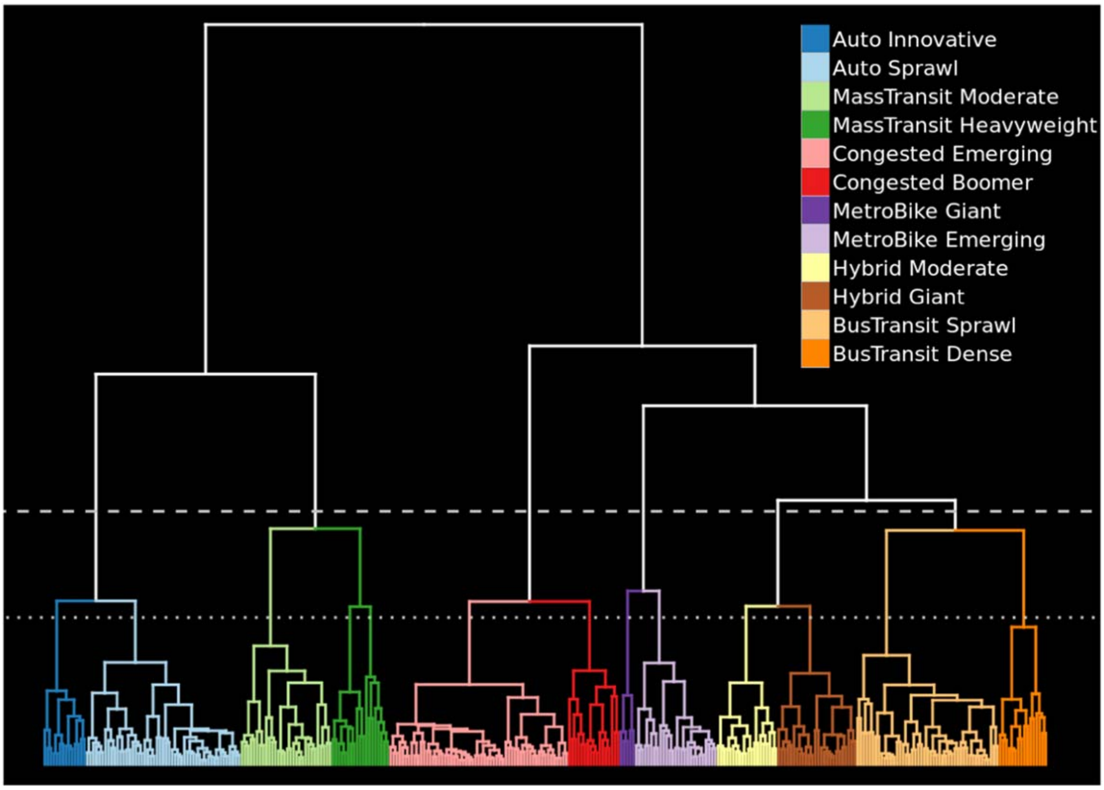

Figure 2. Dendrogram showing classification structure. The dashed line indicates the 6-cluster cut, while the dotted line indicates the 13-cluster cut. The selected result is at the 12-cluster level.

Table 5. Summary of latent urban factors identified.

Factor	Key indicators
Metro propensity	Urban rail/metro (demand, supply, age)
BRT propensity	Bus rapid transit(demand, supply, fares)
Bikeshare propensity	Bikeshare (demand, supply), low cost of living
Development	Wealth, cost of living indices, innovation
Population	Growth, population change
Congestion	Congestion (various metrics), public transit mode share, low car mode share
Sustainability	Bike mode share, street length, safety, efficiency, low congestion
Sprawl	Road deaths, high car mode share, low gas price, CO_2 emissions, street length
Network density	High intersection density, high street density, low street length average, low circuity

3. Results and discussion

3.1. Urban factors

Nine urban factors were discovered from the factor analysis, namely: Metro Propensity, bus rapid transit (BRT) Propensity, Bikeshare Propensity, Development, Sustainability, Population, Congestion, Sprawl and Network Density. These are the dimensions upon which the typologies are defined. The factor descriptions are summarized in table 5. Further, we show the key significant variables and their respective loadings on each factor in figure 3.

3.2. Typology descriptions

The 12 typologies can be grouped into pairs, namely: *Auto*, *BusTransit*, *Congested*, *Hybrid*, *MetroBike* and *MassTransit*. A summary of each typology showing its key features and representative cities is provided in table 6. The factors also enable us to better characterize the typologies, as seen through the spider plot profiles in figure 4. The geographic distribution of the typologies is shown in figure 5.

3.2.1. *Auto Innovative and Auto Sprawl*

The *Auto* cities are modern and highly industrialized, but marked by a history of car-driven development. Thus, both typologies have the lowest mass transit modeshare (figure 6(c)). *Auto Innovative* comprises the subset of relatively dense North American agglomerations with extensive rapid transit systems (e.g. Boston, Toronto, Chicago). Further, it has the highest average Development score (0.82) of all the 12 typologies.

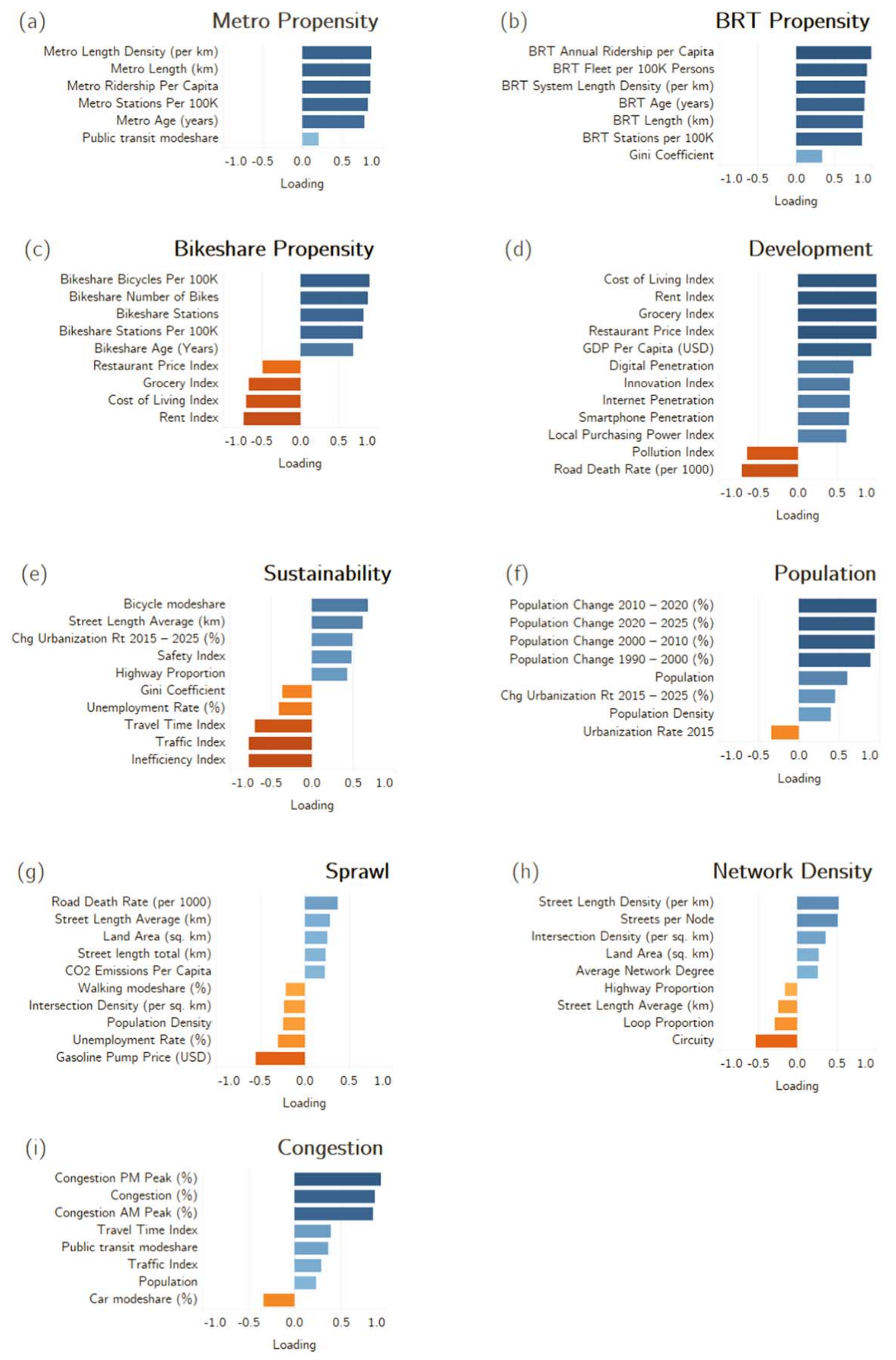

Figure 3. Bar plots showing variables with significant loadings on each urban factor.

Meanwhile, *Auto Sprawl* ranks second highest on the Development factor (0.74), *Auto Sprawl* tellingly scores much lower in Network Density (0.41 compared to 0.58 in *Auto Innovative*). Of all the typologies, *Auto Sprawl* has the highest car mode share and ownership (figure 6(c)), highest emissions (figure 6(b))

Table 6. Summary of the urban typologies and their key cities.

Typology	No. cities	Features; major locations	Key example cities
Auto Innovative	14	Auto-dependent, wealthy, higher transit mode share, metro & population density; U.S., Canada	Washington DC, Boston, Chicago, San Francisco, Toronto
Auto Sprawl	51	Auto-dependent, wealthy, sprawling, lowest transit mode share; U.S., Canada, Middle East	Baltimore, Tampa, Raleigh, Kuwait City
BusTransit Dense	16	Large population, high BRT, fairly congested; South America	Bogota, Rio de Janeiro, Jakarta, Sao Paulo, Tehran
BusTransit Sprawl	47	Lower population, sprawling, fair public transit ; Latin America, Central Asia/Middle East	Mecca, Shiraz, Santa Cruz, Tripoli, Caracas
Congested Boomer	17	Rapid growth, congestion, moderate car mode share; Indian Subcontinent, Africa	Bangalore, Chennai, Delhi, Lagos, Manila
Congested Emerging	59	High growth, lower population, developing; Africa, S. Asia	Kumasi, Phnom-Penh, Port-au-Prince, Lucknow
Hybrid Giant	26	Mix of mode choices, dense networks, high population density; S./E. Europe, E. Asia	Busan, Lisbon, Sapporo, Santiago, Warsaw
Hybrid Moderate	20	Mix of mode choices, lower population; Central America, Middle East	Havana, Johannesburg, Montevideo, Panama City
MetroBike Emerging	27	Metro & bikeshare dominant, highway development, fairly wealthy; China	Ningbo, Zhengzhou, Shenyang, Harbin
MetroBike Giant	5	Metro & bikeshare dominant, large population; wealthy; China	Shenzhen, Guangzhou, Chongqing, Beijing
MassTransit Heavyweight	19	High mass transit usage and metro availability, high bikeshare; fairly high CO_2 emissions; Europe, S.E. Asia	Singapore, Madrid, Seoul, Berlin, London
MassTransit Moderate	30	Equitable, high bikeshare, moderate metro and BRT, low congestion; W. Europe, Israel	Antwerp, Tel Aviv, Turin, Liverpool

and lowest mass transit mode shares (figure 6(c)). Notable examples of this largely North American typology are Baltimore, Indianapolis and St. Louis, including a few cities in the Middle East (e.g. Abu Dhabi, Dubai).

3.2.2. BusTransit Dense and BusTransit Sprawl

These typologies consist of cities with high bus (public transit) usage, along with some of the lowest Development and Sustainability average scores. The *BusTransit Dense* cities have the highest BRT Propensity score (0.56), and they outscore *BusTransit Sprawl* in Population, Congestion and Network Density. However, *BusTransit Sprawl* has the highest Sprawl average (0.72 compared to 0.71 in *Auto Sprawl*). Key examples of *BusTransit Dense* cities are Rio de Janeiro, Jakarta, Tehran and Mexico City. These are highly populated, dense cities with good mass transit systems, particularly with respect to BRT. *BusTransit Sprawl* includes several Latin American cities (Caracas, Puebla, San Salvador), along with others across the world, including Almaty, Cape Town and Isfahan.

3.2.3. Congested Boomer and Congested Emerging

The *Congested Boomer* cities are characterized by joint highest scores in Population and Congestion, compared to other typologies. *Congested Emerging* also features the second highest Congestion average score (0.69). However, both typologies rank the lowest in Development (*Boomer*: 0.14; *Emerging*: 0.08). The *Boomer* cities have the highest Population score (and also the highest average population growth and density figure 7(g)). Notable in this typology are the Indian subcontinental urban centers (e.g. Delhi, Mumbai, Dhaka) along with Lagos, Manila, among others. Remarkably, *Congested Emerging* has the lowest CO_2 emissions per capita and the highest public transit mode share (figures 6(a), (b)). This could be explained by its low levels of industrialization and wealth.

3.2.4. Hybrid Giant and Hybrid Moderate

The *Hybrid* cities are fairly dense, growing urban centers, with solidly average performance across all urban factors. They also have among the highest mass transit mode shares (about 40%) and share similar CO_2 emissions (figure 6(a)) and car usage characteristics (figure 6(b)).

The chief distinguishing factor between the *Hybrid Giant* and *Hybrid Moderate* typologies is Metro Propensity, as *Giant* cities have an average score of 0.37, compared to 0.09 for *Moderate* cities. *Giant* also ranks slightly higher on Congestion and Sprawl compared to *Moderate*. Furthermore, *Giant* cities have an average population of 2.7 million, compared to 2 million for *Moderate* cities. The *Hybrid Moderate* typology is largely represented in South America (e.g. Havana, Cordoba, Panama City) and Central Asia, while the *Hybrid Moderate* cities are chiefly found in Eastern Europe and East Asia (e.g. Daegu, Hiroshima, Sofia).

3.2.5. MetroBike Emerging and MetroBike Giant

These *MetroBike* cities are heavily populated and rapidly urbanizing metropolitan areas in China, where rapid transit and bikeshare systems are widely

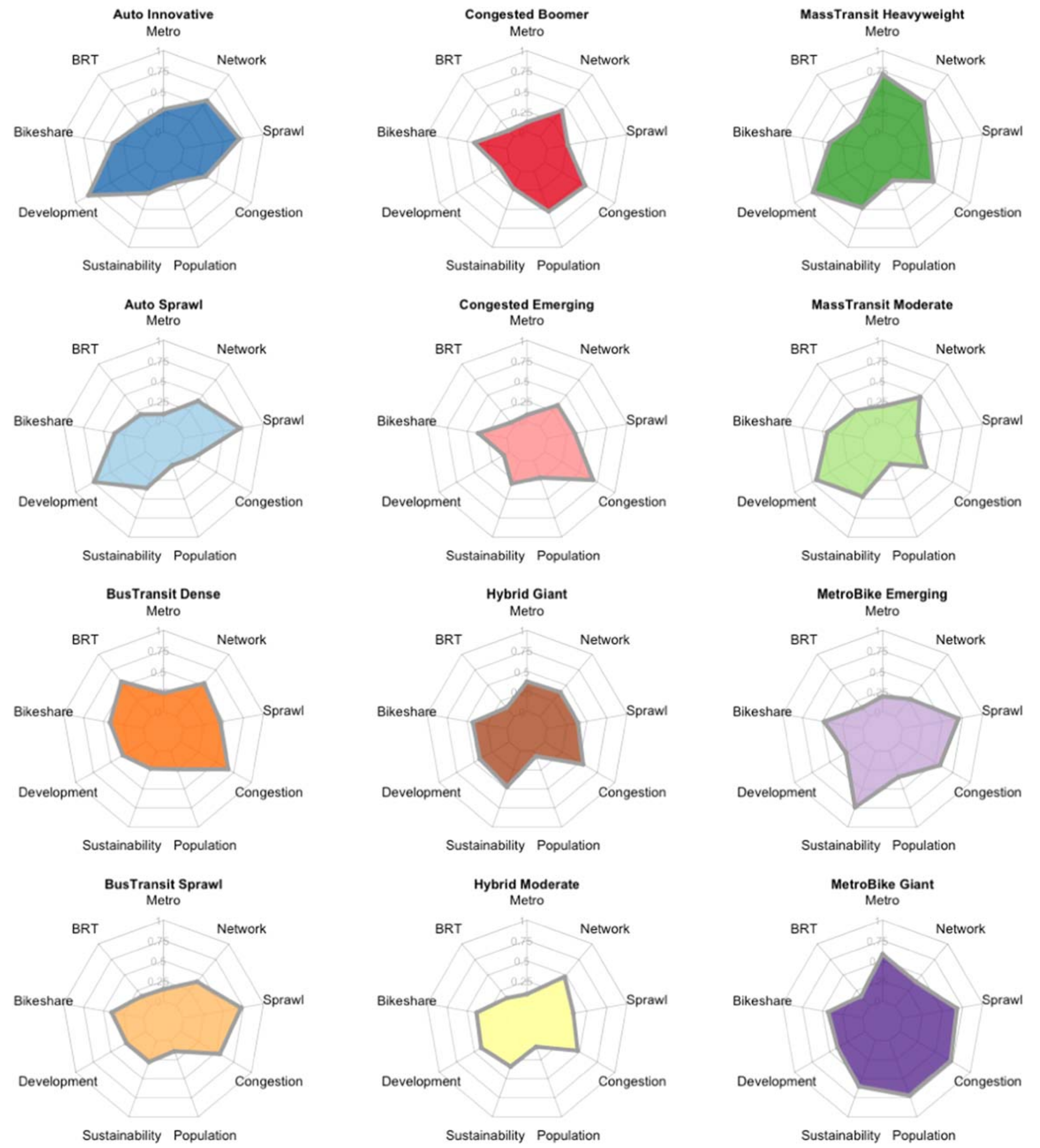

Figure 4. Spider plots indicating normalized factor scores averaged in each typology. Factor names have been simplified for readability. 'Propensity' has been dropped from 'Metro', 'BRT' and 'Bikeshare', while 'Network density' has been shortened to 'Network'.

available. The pace of bikeshare growth in China has been so rapid in recent years that it has posed recent challenges in urban management (Shi *et al* 2018). *MetroBike* cities have the highest Sustainability scores (*Emerging*: 0.74; *Giant*: 0.60) which highlight the prevalence of cycling (*Emerging*: 23%; *Giant*: 17%). We distinguish between the *Emerging* and the *Giant* typologies based on key factors of Metro, Development Population and Congestion, on which the *MetroBike Giant* scores higher (0.58, 0.39, 0.72 and 0.72, respectively) than its *Emerging* counterpart (0.19, 0.27, 0.34 and 0.57, respectively). The *Giant* cities also correspond to the well-known Tier 1 cities in China, such as Shanghai, Beijing and Guangzhou (Li 2007). The *Emerging* cities, which score highest in Bikeshare overall (0.48) include Ningbo, Harbin and Shenyang.

3.2.6. MassTransit Heavyweight and MassTransit Moderate

MassTransit Heavyweight cities have the highest Metro (0.71) and second highest Development (0.74) scores (figure 4). The Development score of this typology is also similar to those of the *Auto* cities. While public transit is fairly high in this typology, CO_2 emissions are the third highest (figures 6(c), (a)). Most member cities are in Europe (London, Berlin, Paris, Oslo, Madrid among others). Singapore, Hong Kong and Tokyo make up the Asian members. The lone members in the Americas are New York City and Vancouver, which are furthest from the centroid of the typology. *MassTransit Moderate* has the second highest BRT (0.26) and Bikeshare (0.48) scores across all typologies. Notably, it has the lowest Population score (0.19) and

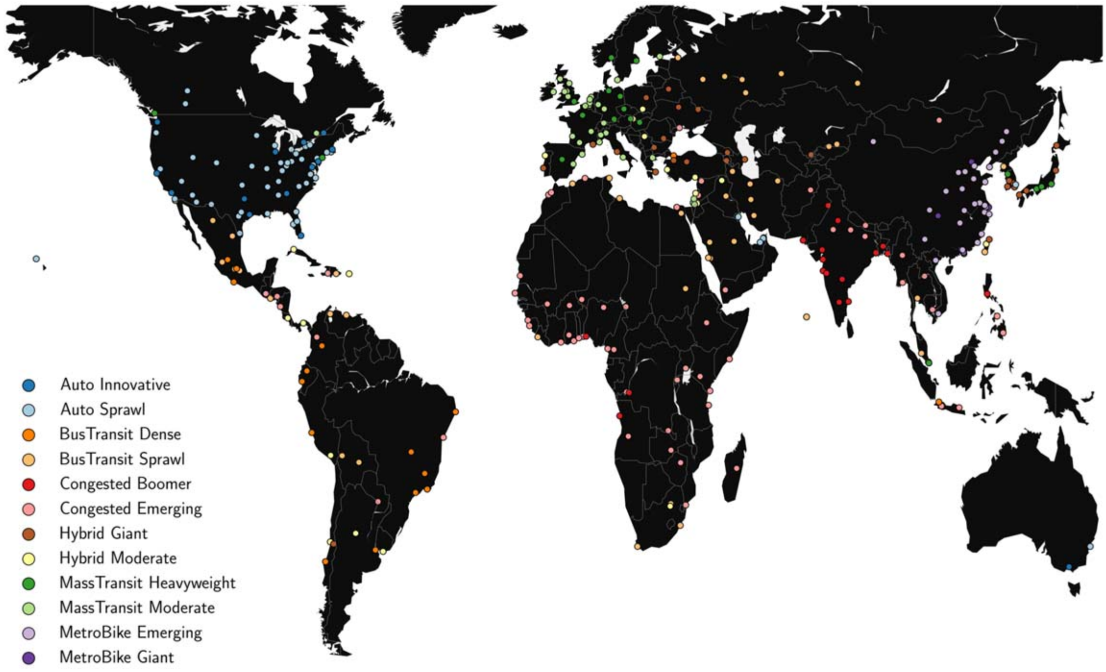

Figure 5. Map of the world indicating cities by typology.

is dominated by European cities (e.g. Antwerp, Brussels, Helsinki), along with major Israeli cities and Ottawa.

3.3. Emissions and urban sustainability

The typology classification provides a unique way of viewing and understanding urban and environmental metrics. Box plots of selected urban and environmental indicators are shown in figure 6 by typology. We see that the *Auto* cities, for example, have the highest car mode share and CO_2 emissions per capita. The *MetroBike* cities stand out with the highest bike mode shares. They also feature the highest highway proportion of all typologies.

Scatter plots of two environmental variables, annual CO_2 emissions per capita and Pollution Index (Numbeo 2017), against selected urban and economic indicators are shown in figure 7. (The Pollution Index is calculated on a scale of 0 to 100, in increasing order of perceived pollution of air, water and the environment.) The data points shown are the average values for each typology. The Vehicles per capita variable was not used in the factor analysis given that it was highly correlated with car usage, and also given that city-level data were not readily available for many cities. From the plots (figures 6(a), (b)), its relationship to CO_2 emissions and Pollution provides validation of the typology classification. For *MetroBike Giant* and *MetroBike Emerging*, the relatively higher emissions/vehicle indicates the impact of the economic and industrial activities that also contribute to their emissions.

On one end of the spectrum are the *Auto Innovative* and *Auto Sprawl* cities (colored light and dark blue) with the highest CO_2 emissions per capita and the lowest Pollution Index. *Congested Boomer* and *Congested Emerging* cities are on the other end of the spectrum with the opposite ranking on those two indicators. Denser cities tend to be more polluted but contribute less to CO_2 emissions per capita. Unsurprisingly the wealthiest cities (those with the highest GDP per capita) contribute the most to per-capita CO_2 emissions. Highway proportion, an indicator of a car-centric mobility infrastructure culture, appears to be a good indicator of CO_2 emissions per capita. The recently developing *MetroBike* cities are an exception here: with the highest highway proportion of around 4%–5%, their contribution to CO_2 emissions is comparable to cities with half that percentage.

We note that the variables discussed here were chosen based on their relevance to environmental sustainability. Further relationships among the variables can be explored across the typologies to yield findings of interest. These could be treated in more depth in a future study.

4. Potential applications

4.1. Prototype city generation for simulation of future mobility

A major application of the typologies presented in this paper is in the creation of prototype cities for large-scale future mobility simulation. The typologies provide a framework for coherently analyzing the impacts of relevant future mobility scenarios across different city types. Thus, we have developed an approach for

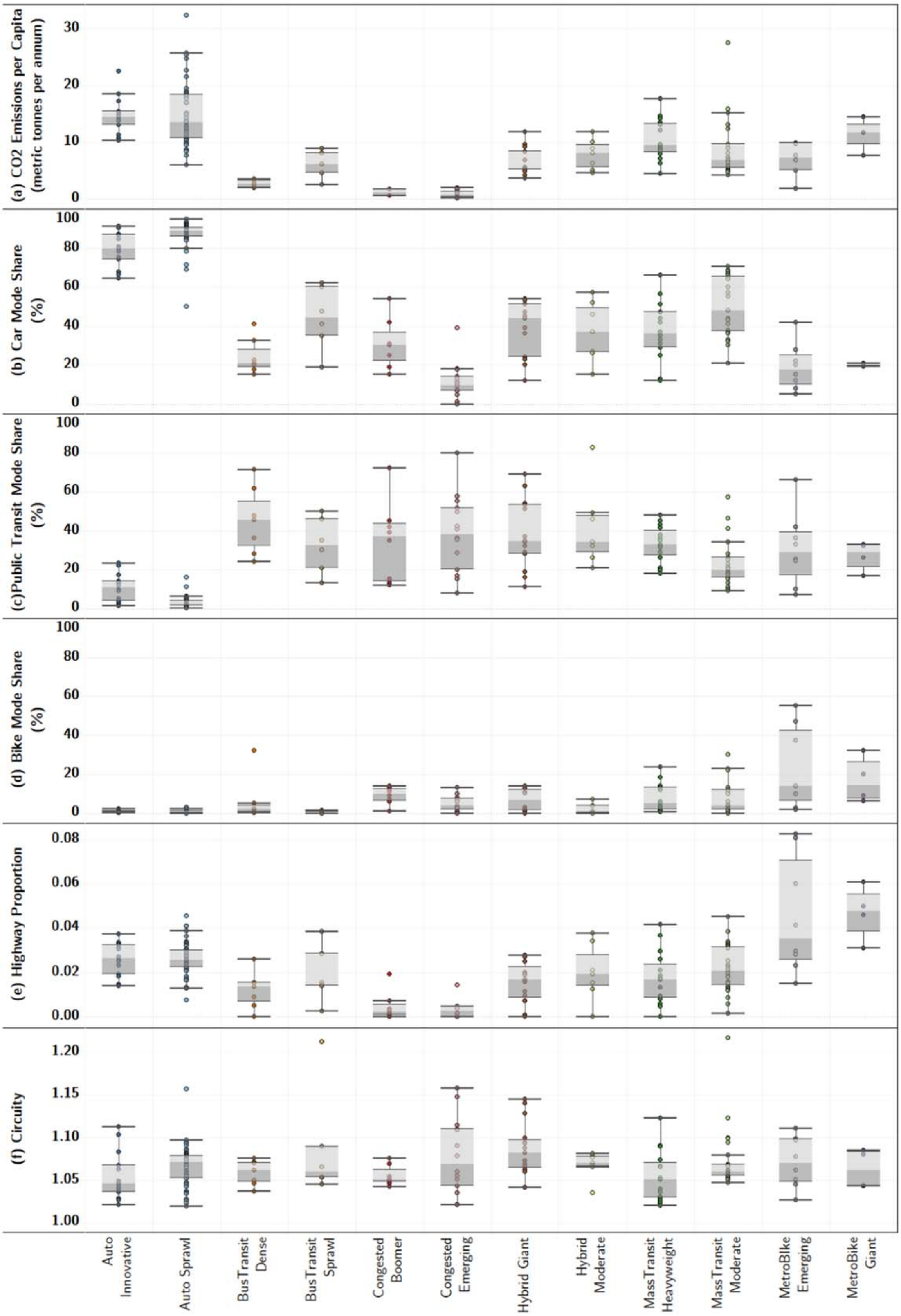

Figure 6. Box plots of selected variables by typology.

high-fidelity large-scale simulation-based analyses of the urban typologies. This hinges on the generation of a prototype city designed to be broadly representative of the cities within a given typology. We generate and validate the prototype city based on population, land-use, demand (activity and mode choice) and supply characteristics, building on earlier efforts to model the Boston metropolitan area in the US (Viegas de Lima *et al* 2018). Using this approach, we have created and simulated prototype cities for *Auto Sprawl* and *Auto Innovative* in a state-of-the-art simulator, SimMobility (Adnan *et al* 2016). These detailed

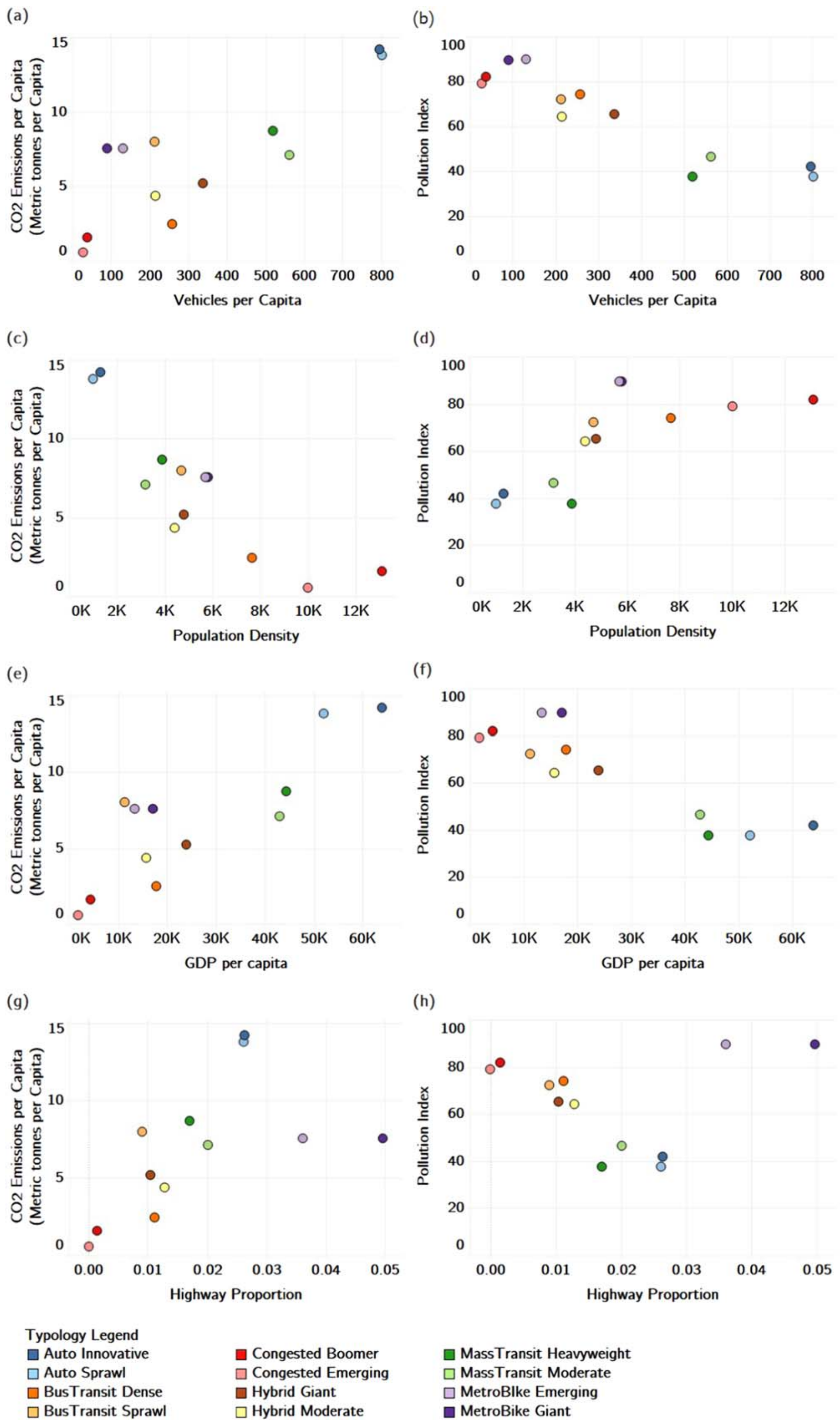

Figure 7. Plots showing relationships between environmental variables (CO_2 emissions per capita and the Numbeo Pollution index) and selected indicators.

simulations have enabled the impact assessment of automated mobility-on-demand scenarios, with results on demand, network and energy outcomes (Oke *et al* 2019).

Agent-based simulations on such prototype cities can yield insights into the broad impacts of future scenarios in these typologies. Considerable time and effort can thus be conserved by using the typology-level evaluations to guide the selection of viable sustainable mobility pathways for more detailed analyses at the city level. Precedents for this sort of application can notably be found in the work of Thomson (1977), where archetype strategies were empirically determined and analyzed for groups across 30 megacities. Fielbaum *et al* (2016) also created a parametric representation of cities as a tool for benchmarking and evaluation of mobility scenarios. In this case, we leverage on the representational power of an agent-based simulator to facilitate typology-relevant scenario explorations.

4.2. Latent class choice model framework for behavioral implications

We have obtained access to a unique behavioral dataset sampled globally from cities across 52 countries (Dalia Research 2017). The survey, conducted by Dalia Research, provides individual characteristic and mobility preference data (car ownership, travel mode, travel time, distance of driving), tendency to purchase and use electronic vehicles (EV), and safety perceptions of autonomous vehicles (AV). Using a latent class choice modeling (LCCM) framework informed by the 9 urban factors discovered, we plan to estimate a model to explain travel behavior to further confirm the validity of the typologies presented in this paper. Moreover, capturing taste heterogeneity in mobility behavior across typology-specific models, the LCCM approach has the added advantage of predicting current, and potentially future, travel-related choices, which are relevant for sustainable urban mobility. The latent urban typology structure is identified by maximizing the total likelihood of observing the choices of individuals from different cities. Ultimately, we can obtain a probabilistic typology profile for each city, which provides richer information compared to a deterministic assignment to only one typology.

5. Conclusion

We have discovered a new set of urban typologies based on 64 urban indicators from 331 cities spanning 124 countries across all continents. The cities represent 40% of the global urban population in 2016, while the indicator data have been compiled from open sources dating back to 2008. Our effort is the most comprehensive mobility-oriented classification to date that also incorporates environmental variables. Using factor-analytic and agglomerative clustering approaches, we identified 9 urban factors and 12 typologies. Each typology captures distinct urban outcomes and serves as a potential testbed for sustainable mobility implementations.

Our results uncover critical typology patterns that will enable researchers to focus future efforts in mitigating environmental concerns. In particular, the *MetroBike* cities in China are the fastest growing, along with the greatest proportion of highways. The *Congested Boomer* typology is also notable for its large population density and congestion problems, yet relatively low CO_2 emissions. Predominantly in North America, the *Auto* cities are exceptional in their car usage, wealth and CO_2 emissions. The *MassTransit Heavyweight* and *MassTransit Moderate* typologies, represent potentially desirable outcomes in sustainable mobility. Further detailed comparative analyses on these typologies can yield valuable insights for urban planners and policymakers.

We have indicated how the typologies can directly impact policy through agent-based simulation of prototype cities. We plan to conduct simulations of alternative mobility scenarios beyond automated mobility-on-demand and on an expanded set of prototype cities, in order to generate insights for optimal policy approaches that cities can adopt to effectively harness new vehicle technologies and mobility services for overall social and environmental benefits. We have also indicated an ongoing extension of the typology discovery effort by incorporating behavior and perceptions using a recent survey. With the aforementioned LCCM approach, we expect to further enhance our typology specifications and provide insights into the likelihood of EV and AV adoption across various urban typologies.

Our data and results have been made publicly available at http://its.mit.edu/typologies. We expect that these will be valuable to researchers and planners as a foundation for further research, as well as aid policy efforts in reducing the environmental impacts of urban mobility for a sustainable future.

Acknowledgments

This work was supported by the Massachussets Institute of Technology Energy Initiative's Mobility of the Future study. We also thank Sean Hua, Joseph Noszek, Jin (Jasmine) Zhou and Eytan Gross (Massachusetts Institute of Technology) for their contributions in the early stages of data compilation and analyses. Finally, the authors would like to thank the reviewers for their feedback which improved the quality of this manuscript.

ORCID iDs

Jimi B Oke https://orcid.org/0000-0001-6610-445X

References

Adnan M, Pereira F C, Lima Azevedo C M, Basak K, Lovric M, Raveau S, Zhu Y, Ferreira J, Zegras C and Ben-Akiva M E 2016 SimMobility: a multi-scale integrated agent-based simulation platform *Transportation Research Board 95th Annual Meeting* number 16-2691 (https://trid.trb.org/view/1393000)

Basu R, Araldo A, Akkinepally A P, Nahmias Biran B H, Basak K, Seshadri R, Deshmukh N, Kumar N, Azevedo C L and Ben-Akiva M 2018 Automated mobility-on-demand vs. mass transit: a multi-modal activity-driven agent-based simulation approach *Transportation Res. Rec.* **2672** 608–18

Bland D and Hernandez M 2016 Urban legend: China's tiered city system explained 201 (http://multimedia.scmp.com/2016/cities/)(Accessed: 30 December)

Blashfield R K 1976 Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods *Psychol. Bull.* **83** 377

Boeing G 2017 OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks *Comput. Environ. Urban Syst.* **65** 126–39

Bruce G D and Witt R E 1971 Developing empirically derived city typologies: an application of cluster analysis *Sociol. Q.* **12** 238–46

Celimli Aksoy S 2017 A comparison of geomin versus target rotation criteria in exploratory factor analysis with correlated factors and large and complex pattern matrices *PhD Thesis* (Coral Gables, FL: University of Miami)

Cham H, Reshetnyak E, Rosenfeld B and Breitbart W 2017 Full information maximum likelihood estimation for latent variable interactions with incomplete indicators *Multivariate Behav. Res.* **52** 12–30

Charrad M, Ghazzali N, Boiteau V and Niknafs A 2014 NbClust: an R package for determining the relevant number of clusters in a data set *J. Stat. Softw., Artic.* **61** 1–36

Creutzig F, Baiocchi G, Bierkandt R, Pichler P-P and Seto K C 2015 Global typology of urban energy use and potentials for an urbanization mitigation wedge *Proc. Natl Acad. Sci.* **112** 6283–8

Dalia Research 2017 How the world moves: the future of mobility and global transport (https://daliaresearch.com/wp-content/uploads/2017/05/27-3-2017_pressrel_future_mobility-1.pdf)

Demographia 2017 (http://demographia.com/db-worldua.pdf) (Accessed: 3 January 2017)

Dobbs R, Smit S, Remes J, Manyika J, Roxburgh C and Restrepo A 2011 Urban world: mapping the economic power of cities *Report* no. 62 McKinsey Global Institute

Ferreira L and Hitchcock D B 2009 A comparison of hierarchical methods for clustering functional data *Commun. Stat.—Simul. Comput.* **38** 1925–49

Fielbaum A, Jara-Diaz S and Gschwender A 2016 A parametric description of cities for the normative analysis of transport systems *Netw. Spat. Econ.* **17** 343–65

Global BRTData 2017 (http://brtdata.org)(Accessed: 10 January 2017)

Global Petrol Prices 2017 (http://globalpetrolprices.com/gasoline_prices)(Accessed: 2 February 2017)

Global Urban Indicators Database 2015 Global Urban Observatory Unit, UN-Habitat Accessed on: 18 October 2016

Hadden J and Borgatta E 1965 *American Cities: Their Social Characteristics* (Chicago: Rand McNally & Co.)

Hands S and Everitt B 1987 A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques *Multivariate Behav. Res.* **22** 235–43

Harris C D 1943 A functional classification of cities in the United States *Geogr. Rev.* **33** 86–99

Hattori M, Zhang G and Preacher K J 2017 Multiple local solutions and geomin rotation *Multivariate Behav. Res.* **52** 720–31

Hayashi K, Bentler P M and Yuan K-H 2007 On the likelihood ratio test for the number of factors in exploratory factor analysis *Struct. Equ. Model.: Multidiscip. J.* **14** 505–26

Hirose K, Kim S, Kano Y, Imada M, Yoshida M and Matsuo M 2016 Full information maximum likelihood estimation in factor analysis with a large number of missing values *J. Stat. Comput. Simul.* **86** 91–104

Huang J, Lu X and Sellers J M 2007 A global comparative analysis of urban form: applying spatial metrics and remote sensing *Landscape Urban Plan.* **82** 184–97

IEA 2017 CO_2 Emissions from Fuel Combustion 2017 *Technical Report* (Paris: OECD Publishing) (https://doi.org/10.1787/co2_fuel-2017-en)

Innovation Cities Index 2017 (http://innovation-cities.com/innovation-cities-index-2016-2017-global) (Accessed: 8 June 2017)

Internet World Stats 2017 (http://internetworldstats.com) (Accessed: 10 August 2017)

ITF 2017 ITF Transport Outlook 2017 (Summary) (Paris: OECD Publishing) (https://doi.org/10.1787/e979b24d-en)

Jain A K 2010 Data clustering: 50 years beyond k-means *Pattern Recognit. Lett.* **31** 651–66 Award winning papers from the 19th Int. Conf. on Pattern Recognition (ICPR)XIX International Conf. in Pattern Recognition (ICPR)

Jain A K, Murty M N and Flynn P J 1999 Data clustering: a review *ACM Comput. Surv.* **31** 264–323

Jöreskog K G 1969 A general approach to confirmatory maximum likelihood factor analysis *Psychometrika* **34** 183–202

Kamakura W A and Wedel M 2001 Exploratory Tobit factor analysis for multivariate censored data *Multivariate Behav. Res.* **36** 53–82

Kuiper F K and Fisher L 1975 391: a Monte Carlo comparison of six clustering procedures *Biometrics* **31** 777–83

Le Néchet F 2012 Urban spatial structure, daily mobility and energy consumption: a study of 34 European cities *Cybergeo: Eur. J. Geogr.* 580

Li C 2007 The leadership of China's four major cities: a study of municipal party standing committees *China Leadership Monitor* **21** 1–19

Lorenzo-Seva U 2003 A factor simplicity index *Psychometrika* **68** 49–60

Martin M K, Carley K M and Altman N 2008 300 Cities—an exploration in characterizing US cities *CASOS Technical Report* CMU-ISR-08-122 Institute for Software Research, Carnegie Mellon University, Pittsburgh, PA 15213

McIntosh J, Trubka R, Kenworthy J and Newman P 2014 The role of urban form and transit in city car dependence: analysis of 26 global cities from 1960 to 2000 *Transp. Res.* D **33** 95–110

Milligan G W and Cooper M C 1988 A study of standardization of variables in cluster analysis *J. Classif.* **5** 181–204

Murtagh F and Legendre P 2014 Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* **31** 274–95

Muthen B O 1989 Tobit factor analysis *Br. J. Math. Stat. Psychol.* **42** 241–50

Muthén L and Muthén B 2017 *Mplus User's Guide* 8th edn (Los Angeles, CA: Muthén & Muthén)

Numbeo 2017 (https://numbeo.com/common) (Accessed: 9 July, 2017)

Oke J B, Akkinepally A, Chen S, Xie Y, Aboutaleb Y M, Azevedo C L, Zegras P C, Ferreira J and Ben-Akiva M E 2019 Evaluating systemic effects of automated on-demand services through large-scale, agent-based simulation of auto-dependent, prototype cities *Transp. Res.* (accepted)

OpenStreetMap 2017 (https://openstreetmap.org)(Accessed: 21 September, 2017)

Poushter J 2016 Pew Research Center Smartphone ownership and internet usage continues to climb in emerging economies (https://www.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf)

Price D O 1942 Factor analysis in the study of metropolitan centers *Soc. Forces* **20** 449–55

Priester R, Kenworthy J and Wulfhorst G 2013 The diversity of megacities worldwide: challenges for the future of mobility *Megacity Mobility Culture* (*Lecture Notes in Mobility*) ed Institute for Mobility Research (Berlin: Springer) pp 23–54

Satorra A and Bentler P M 2001 A scaled difference chi-square test statistic for moment structure analysis *Psychometrika* **66** 507–14

Shi J-G, Si H, Wu G, Su Y and Lan J 2018 Critical factors to achieve dockless bike-sharing sustainability in China: a stakeholder-oriented network perspective *Sustainability* **10** 2090

Thomson J M 1977 *Great Cities and Their Traffic* (London: Victor Gollancz Ltd)

TomTom 2016 (https://tomtom.com/en_gb/trafficindex/) (Accessed: 11 November, 2016)

UN-Habitat 2015 (http://urbandata.unhabitat.org) (Accessed: 17 October, 2016)

U.S. DOT 2017 Smart city challenge: lessons for building cities of the future *Technical report* U.S. Department of Transportation, Washington, D.C. (https://transportation.gov/policy-initiatives/smartcity/smart-city-challenge-lessons-building-cities-future)

Viegas de Lima I, Danaf M, Akkinepally A, De Azevedo C L and Ben-Akiva M 2018 Modeling framework and implementation of activity- and agent-based simulation: an application to the Greater Boston area *Transp. Res. Rec.* **2672** 146–57

World Health Organization 2013 (http://apps.who.int/gho/data/node.main.A997) (Accessed: 22 October, 2016)

Yates A 1988 *Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis* (Albany, NY: State University of New York Press)

Zegras P C and Gakenheimer R 2005 Sustainable urban mobility: exploring the role of the built environment *Doctoral Dissertation* (Cambridge, MA: Massachusetts Institute of Technology)