Urban form influences travel distances, car ownership, and mode choice: Evidence from 19 European cities

Aneeque Javaid
Mercator Research Institute on Global Commons and Climate Change

Nikola Milojevic-Dupont
Mercator Research Institute on Global Commons and Climate Change

Florian Nachtigall
Technische Universität Berlin

Felix Wagner
Mercator Research Institute on Global Commons and Climate Change

Felix Creutzig
Mercator Research Institute on Global Commons and Climate Change

Peter Berrill (✉ peter.berrill@aya.yale.edu)
Technische Universität Berlin  https://orcid.org/0000-0003-1614-3885

Article

Keywords:

Posted Date: May 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2924076/v1

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

Steady growth in global greenhouse gas emissions from transport is driven by growing demand for car travel. Cities hold large potential to reduce energy demand and emissions from mobility through encouraging shorter travel distances and sustainable travel modes. In European cities however, personal cars still dominate travel, facilitating continued growth of transport emissions and having negative implications for numerous other dimensions of sustainability. A growing body of research investigates linkages between urban form and mobility, mostly using aggregate data in multiple cities, or disaggregated data for individual cities. Here, we compare urban travel patterns and influences of urban form at spatially disaggregated scale across nineteen cities in four European countries using statistically advanced methods. We enrich travel survey data with metrics describing local urban form. We compare car ownership and travel patterns across cities and use supervised machine learning to explore influences of urban form and other features on mode choice, car ownership, and trip distances. Residential proximity to the city center is the greatest enabler of sustainable urban mobility. Future residential development should be concentrated near to urban centers. Overall city size is important, as occupants of small and medium-sized cities have higher car ownership and use than large cities, motivating increased attention on sustainable mobility transitions outside of large cities. We highlight targeted solutions to increase access to sustainable mobility for certain population groups, and for longer urban trips. Our results confirm that urban planning is a key instrument for increasing sustainability of land transport.

Introduction

Global greenhouse gas (GHG) emissions from transport grew 73% between 1990 and 2019, and constituted 27% of 2019 energy-related emissions (Minx et al., 2021). Emissions from transport are challenging to mitigate, and the sector has been referred to as a ‘roadblock’ to climate change mitigation (Creutzig et al., 2015). But not all transport emissions are equally difficult to mitigate. Emissions arising from urban mobility, responsible for about 40% of transport emissions (Creutzig et al., 2016), can be more easily reduced compared to other transport emissions, for two major reasons. First, viable alternative modes of transportation with low energy and emissions per passenger-km such as active travel and public transport are more widely accessible in cities, which usually leads to lower car mode shares in urban areas (Pucher & Renne, 2005; X. Tao et al., 2019; Ton et al., 2020). Second, per capita travel demand tends to be lower for urban dwellers, as distances between origins and destinations are shorter (Holz-Rau et al., 2014; Pucher & Renne, 2005). Emissions reductions through mode choice shifts come with strong health co-benefits from reduced air pollution exposure and increased active travel (Stevenson et al., 2016; Woodcock et al., 2009), as shorter trips are more likely to involve active travel modes. Higher population concentrations lead to greater health benefits from reduced air pollution in cities (Apte et al., 2012), which affects mental as well as physical health (Cao et al., 2023). Despite the compelling health and environmental benefits of lowering car travel, and the greater feasibility of alternative modes in urban contexts, cars still account for the majority of distance traveled in cities worldwide (Verbavatz &
Barthelemy, 2019). It remains unresolved how morphological features of cities can be optimized in order to accelerate a reduction in emissions from urban travel.

Urban form is a term for describing physical characteristics of urban built environments, and the spatial distribution of those characteristics within a city. Common metrics used to define urban form include density (population, built-up, employment), destination accessibility, street design, and land-use diversity (Ewing & Cervero, 2010). Urban form and infrastructure are cited as key factors for mode choice and emissions from urban mobility (Javaid et al., 2020), with an estimated emission mitigation potential of around 2 Gt CO$_2$/yr (Creutzig et al., 2022). There is a depth of literature which investigates connections between urban form and sustainable mobility outcomes, including travel distances (T. Tao & Næss, 2022; Wagner et al., 2022), mode choice (Kim, 2021; F. Wang & Ross, 2018), transit ridership (Ding et al., 2019), car ownership (Sabouri et al., 2021), and mobility emissions (Wu et al., 2019). The influence of urban form on sustainable mobility is heterogeneous within cities (Wagner et al., 2022), as well as between cities (Ewing et al., 2018), but this multidimensional heterogeneity is rarely studied within a single framework, as very few studies combine spatially disaggregated (within city) data with a broad geographic scope. To our knowledge, Sabouri et al. ‘s (2021) study of car ownership in metro areas of the US is the only study with both a large number of cities and disaggregated indicators of urban form.

Here, we assemble and harmonize urban mobility survey data from nineteen cities in France, Germany, Spain, and Austria. These cities include the capital and largest cities in each country, and a mix of small and medium-sized cities in France and Germany. The combined data describes 670,000 trips by 250,000 individuals from 100,000 households. We enrich the harmonized mobility data with descriptions of urban form generated for this study using openly available data, at the level of local administrative units used in each survey (postcode or higher resolution). Urban form variables gathered include population and built-up density, distance to city center and subcenters, street network characteristics, land-use shares, and time to transit when available (Table 1). We first compare and discuss high-level statistics and patterns between cities. We then use explainable machine learning methods to specify gradient boosting decision tree (GBDT) models that predict three outcome variables – average trip distance, car ownership, and trip mode choice. These three outcome variables are chosen because they together mechanistically determine the energy demand, GHG emissions, and other environmental impacts from urban travel. Understanding the variables which have importance to these outcomes can give policymakers and planners actionable insights for reducing emissions from urban mobility. These insights are particularly relevant to ‘avoid’ (lower travel distance) and ‘shift’ (more sustainable mode choice) demand-side strategies (Creutzig et al., 2018). ‘Improve’ strategies which cover transitioning to low-emission energy sources for motorized travel are a further key pillar in decarbonizing all transport emissions but are beyond the scope of this analysis.

The EU has ambitious climate targets including a 55% reduction in GHG emissions between 2005 and 2030, and zero net emissions by 2050. Progress since 2005 has been made, predominantly from cleaner electricity production, but transport remains the only sector in the EU where emissions continue to grow (European Environment Agency, 2022). Swift action is required to address transport emissions using all
available measures. Our investigation suggests a large direct influence of urban form on trip distances, a moderate influence of urban form on car ownership, and a combination of large indirect and smaller direct influences of urban form on mode choice. Car ownership, a crucial factor for continuing the unsustainable status-quo in urban mobility, is closely correlated with income and demographic variables, but also has marked correlations with urban form, particularly in large cities. Mode choice is most strongly correlated with trip distance, followed by car ownership, demographics, and urban form. The results of this study have important implications for urban areas of all sizes in Europe and other developed world regions.

Results

Comparison of mode shares, car use, and car ownership

Substantial differences in mode shares and car use are evident across the cities studied, and city-level population density appears to account for some of this variation. Car mode shares range from 39% in Paris and Berlin to 84% in Clermont (Fig. 1) and have a clear negative linear correlation (Pearson's $r$) of -0.65 with aggregate city population density. Transit shares range from 9% in Clermont to 52% in Paris and have a strong positive correlation of 0.76 with population density. Population density does not explain all of the variation in mode shares – local contexts including prices, policies, mode-specific infrastructure, and culturally rooted attitudes likely all play some role (Javaid et al., 2020). Wien and Potsdam, for instance, have higher transit shares than the mode share-density trend would suggest, while variations in bike mode shares appear to correlate mainly with country (German cities have far higher biking shares than non-German cities).

The non-linear decline of car use with increasing city-level population was famously demonstrated by Newman and Kenworthy (Newman & Kenworthy, 1989), although more recent interpretations point out that this relationship may weaken as spatial resolution increases (i.e. with disaggregated city data) (Ewing et al., 2018). At postcode (or similar) level in the cities examined here, the decline of car travel and mode share with increasing density appears non-linear in most cases, most notably in French cities outside of Paris. Only in Madrid is the decline close to linear (Fig. S3). Car travel and mode share increase approximately linearly with distance to city center, but the rate of increase is not constant across cities - slopes range from 0.5 to 1.2. In Paris for instance, a simple linear model estimates that living 1km further from the city center correlates with a 0.5km increase in daily car travel. Slopes are highest (>1) in French cities outside of Paris. For mode share, simple linear models estimate that living an additional 1km from the city center correlates with a ~2% increase in car mode share in most cities, except for Madrid, where a weaker increase of 1% is observed (Fig. S3).

Household car ownership rates follow a distinct sigmoid-shape relationship with household income across all eleven cities where income data was collected (all German cities plus Paris, Clermont, and Toulouse), but the ownership saturation rates differ across country and city sizes (Fig. 2a). Car ownership is a crucial variable for understanding and predicting travel patterns (Sabouri et al., 2021), and a
Gompertz sigmoid function has previously been demonstrated for the relationship between car ownership and income at country level (Dargay et al., 2007). We also observe sigmoid relationships between mode share with income (Fig. 2b). Car ownership and mode share tend to saturate at household incomes of around 2,500-3,000 €/month. In the two large cities with income data (Berlin and Paris), car ownership rates are substantially higher in Paris, but car mode share is almost identical. City size appears to play a role, as car ownership and particularly mode share saturate at considerably lower levels in large cities (Berlin and Paris) compared to mid-size and small cities. Between countries, car ownership and mode share at each income level are higher for French cities compared to similarly sized German cities, although the small number of cities involved in this comparison hinders the generalizability of this finding. At incomes of 2,500 €/month, 97% of households in Clermont own a car, and 89% of p-km traveled are by car. At similar incomes in Magdeburg (comparable to Clermont in size and density), 80% of households own a car, and cars fulfill 66% of p-km traveled.

Overall model summaries

To summarize variable importance for each model and region, Figure 3 illustrates the variables with the largest contribution to the model prediction for the models of trip distance, car ownership and mode choice. Trip distance (averaged by residential administrative unit) is strongly correlated with distance to city center, share of commuting trips, and at least one of population, street intersection, or built-up density. Car ownership is most strongly associated with income status, household demographics, and urban form. Distance to center is the most important urban form feature for car ownership and has a higher importance in the large cities of Berlin and Paris than the smaller cities. Associations between urban form and both trip distance and car ownership are slightly weaker in smaller cities, while car ownership shows a stronger response to increases in income in French cities. Mode choice, at the individual trip level, is most strongly associated with our two other outcome variables: trip distance and car ownership. Car ownership has lower importance on mode choice in regions (e.g. rest of France) where household car ownership was ubiquitous. Individual demographics (age and sex), trip characteristics (if it is a ‘companion’ trip with the purpose of accompanying other householders to a destination), and urban form features (most notably distance to center and population density) are all of moderate importance to trip mode choice. These summary insights regarding variable importance in each model are combined into a graph showing the strongest associations of urban form and other explanatory variable groups on our three outcome variables, and ultimately GHG emissions of urban mobility (Fig. S14).

We now describe in more detail the results for each model. Trip distance is modeled using a GBDT regression model, and car ownership and trip mode choice are modeled using GBDT classifier models, using the xgboost library in python. The influence of urban form and other feature variables in each model is demonstrated using Shapley additive explanation (SHAP) values, developed by Lundberg and colleagues (Lundberg et al., 2020). For tree ensemble machine learning methods, this algorithm assigns each model feature an importance value signifying how it influences each individual prediction. We note that feature importance values do not provide true causal effect size estimates, especially in the case of
multicollinearity. When multiple features are correlated and convey similar information, tree-based models focus on one feature, resulting in SHAP values that, relative to the true causal effect, can overestimate the importance of some features and underestimate others. Further details on model specification and validation is given in the Methods section. For ease of presentation, we combine model German cities except Berlin and French cities except Paris together as ‘Germany, other’, and ‘France, other’, respectively.

**Trip distance**

In our trip distance model, we examine how urban form and summary statistics of household income and trip purpose influence average trip distance per spatial unit. To mediate the considerable variation in distances for individual trips, we aggregate and average trip distances based on residential location, as that is one of the main determinants of overall travel demand (Næss et al., 2018; Srinivasan & Ferreira, 2002). This means that our model explores how urban form at the residential location affects all downstream trips, including those that do not start or finish at home. Figure 4 maps the combined influence of urban form variables on average trip distance for the four largest cities in our study, Berlin, Madrid, Paris, and Wien. Blank areas indicate regions where no survey respondents live. This is most visible for Madrid, where large non-residential areas (including parks and industrial areas) exist relatively close to the dense urban core.

Figure 4 aggregates together influences of densities, distance to center and subcenters, street network design and land use shares. The combined urban form results illustrate quite clearly the importance of residential proximity to the city center. Urban form characteristics induce trip distances which can be multiple km longer or shorter than city average. The total difference in average trip distances within cities attributed to local urban form effects is up to 4-5 km in Berlin, Madrid, and Wien, and 2 km in Paris. In Wien for instance, influences of urban form on average trip distance ranges from -2 km (near the center) to +3 km (on the outskirts). In Madrid urban form induces increases in average trip distance of up to 4 km in a few peripheral regions, and reductions of approximately 1 km in the urban core. Among urban form features, distance to the city center has the highest importance in all cities (Fig. S10). This is also true for combined models for the rest of Germany (excluding Berlin) and the rest of France (excluding Paris).

**Car Ownership**

Car ownership is modeled as a function of household income, demographic characteristics (household size and max householder age), and urban form features, for the cities whose surveys include income data. Among urban form features, distance to the city center usually has the highest importance. It is the first and second most important of all features (including non-urban form features) in Berlin and Paris, respectively, the third most important feature in the rest of France, and fifth most important feature in the rest of Germany (Fig. S12). Non-linear effects and thresholds of increasing car ownership with increased distance to center can be observed (Fig. 5). In Berlin a clear threshold of increasing likelihood of car
ownership is observed between 6-8 km from the city center. In Paris, there is a substantial increase in likelihood of car ownership at 5 km from the city center, and in the rest of France (Clermont and Toulouse) a steady increase in likelihood of car ownership is seen between 5-7 km from the city center. In the rest of Germany, a smaller increase in car ownership is observed between 4-5 km from the city center.

In keeping with previous literature (Dargay et al., 2007; Heinonen et al., 2021; Sabouri et al., 2021), income is identified as a primary determinant of car ownership. Here, we find higher importance of income in French cities than German cities (Fig. 3). Higher household size increases the likelihood of car ownership across cities - with the highest increase in likelihood for car ownership is observed when going from household size 1 to 2 (Fig. S13). Although income data were not available to model car ownership at household level in Madrid and Wien, the other large cities in our sample, maps of car ownership suggest that distance to city center has a lower influence on car ownership in Madrid and Wien (Fig. S8). Citywide household car ownership rates are 57% in Berlin and 64-72% in other German cities, 68% in Paris and 75-85% in other French cities, 75% in Madrid, and 76% Wien.

In Berlin and Paris, but not the rest of Germany or France, younger households have lower likelihood of car ownership (Fig. S12). This suggests an interaction of age and city size, where younger households in bigger cities have lower propensity to own a car. One possible explanation for this is increased access to car sharing in (larger) cities (Giesel & Nobis, 2016). Cultural effects, including preferences for living car-free may also play a role (Lee-Gosselin, 2017). Car ownership decreases with increasing population density in all cases. Non-linear effects can be seen (Fig. S13), and although specific thresholds differ across cities, population densities lower than 50 per/ha are generally associated with high probabilities for car ownership. Built-up density can be an important predictor of car ownership, but the effect appears to be relevant only at the lowest levels of built-up density, where car ownership is highly probable (Fig. S13).

Mode choice

Trip distance is the most important feature for predicting mode choice in all cities (Fig. 3). For car mode choice specifically, car ownership is often most important, while companion trip purpose has a consistently high ranking (Fig. 6). With the exception of Berlin, companion trips (which usually involve accompanying children to or from a destination) are very likely to be made by car, especially if the trip is longer than 1 km (Fig. 6, Fig. S6). Age is an important predictor of mode choice, but the main influence of age is a lower likelihood of traveling by car between ages of approximately 10-20, and a higher likelihood of traveling by car above 70, with little difference in car mode choice between the ages of 20-70. Longer trips are more likely to be taken by car or transit, while shorter trips are more likely to be made by foot or bike (Fig. S11). High importance of trip distance and car ownership have been identified in previous studies linking urban form and mode choice (Hagenauer & Helbich, 2017; Liu et al., 2021; F. Wang & Ross, 2018), but the significance of companion trips for mode choice has not been so explicitly demonstrated, although higher car dependence of companion trips has been noted (Leroutier & Quirion, 2022). Distance
to the city center is the most highly ranked urban form predictor of mode choice in Paris, Madrid, and the rest of France (Fig. 6), and it can be important in deciding mode choice for longer trips - living further from the city center makes traveling by transit less likely and by car more likely. Population density is of moderate importance across all cities, mainly for traveling by car. Female travelers are more likely to travel by transit than males and can be less likely to engage in active travel modes, especially in Madrid and the rest of France (Fig S11). Urban form overall makes moderate direct contributions to mode choice prediction, but the indirect influence of urban form on mode choice, through the mediating variables of trip distance and car ownership, seems considerably larger (Fig. S14).

Discussion

Three main recommendations arise from our analysis, with relevance to local and national decision makers. The first recommendation is to concentrate future residential developments and population growth close to city centers, preferably within 5km of the city center and in areas exceeding 50 per/ha, to leverage the non-linear effects associated with these urban form features (Fig. 5, Fig S10, Fig S13). To increase local acceptability and pursue an additional sustainability co-benefit, the added real estate value of densification can be used to finance retrofits of low efficiency older buildings (Ferrante et al., 2020). Distance to center has the greatest importance in models of trip distance, and has considerable importance for household car ownership, particularly in Berlin and Paris. The direct importance of distance to center on mode choice is moderate, but mode choice is mainly influenced by trip distance and car ownership, and so the indirect importance of residential distance to center on mode choice appears substantial. Increasing the share of population living in proximity to city centers can reduce car ownership rates, trip distances, and car mode shares, assuming the correlations identified here by explainable machine learning methods have some basis in underlying causal mechanisms. Various approaches have recently been developed to explore causality using qualitative and graph-based methods (Runge et al., 2019; T. Tao & Næss, 2022; J. Wang et al., 2021), and extending the approach of current study within a causality framework is a promising area for future research. A causal relationship between distance to city center and trip distance is a common assumption, and is a core tenet of urban economics models where workers commute to the center (Creutzig, 2014; Fujita, 1989).

The second recommendation based on our results is to implement measures which reduce the car mode share for longer trips, which are mostly carried out either by car or transit (Fig. 6, Fig S5, Fig S11). Increasing frequency and coverage to transit for people living further from city centers, and increasing the costs of private car use are two promising strategies in this regard (Axsen et al., 2020; Ingvardson & Nielsen, 2018). Increasing transit frequency may be achievable in a relatively short timeframe, but increasing coverage is a longer-term effort. A number of instruments can increase costs of private car use, such as increased prices for fuel, parking, and/or entering cities with cars (e.g. through congestion charging or low-emission zones). These instruments are likely to be politically unpopular, at least initially, and can have undesirable equity implications, as lower income residents can be more car dependent and have lower access to sustainable modes (Mattioli, 2021; Wagner et al., 2022). However, evidence points to the necessity of some push factors in combination with pull factors to replace car use with more
sustainable alternatives (Axsen et al., 2020; Liotta et al., 2022; Xiao et al., 2022). An extension of the current study could identify and mitigate risks of disproportionately targeting lower income households with the costs of sustainable mobility transitions. For instance, if implementing a congestion charge or other measure targeting car use, areas with low income and high car dependency (as illustrated by Wagner et al. (2022) for the case of Berlin) might require intermediary extensions to bus services until longer term solutions are ready.

Our third recommendation, expanding upon the second, calls for a focus on particular population groups to target solutions for reducing car use for trips of all distances. Companion trips (accompanying children) are more likely to be done by car (Fig. 6), especially for trips over 1 km (Fig. S6). 'Pull'-oriented solutions for increased transit use for companion trips may involve lower costs for traveling with children, and increased network connectivity to facilitate more complex trip chaining (Hensher & Reyes, 2000). Shared mobility, including ride-sourcing and car sharing, could displace private car use for family trips and lessen the propensity for car ownership among households with children, although evidence to date is inconclusive regarding engagement of households with children in shared mobility (Amirnazmiafshar & Diana, 2022), and mixed regarding car shedding due to engagement in shared mobility (Diao et al., 2021; Giesel & Nobis, 2016). To increase bike mode share for population groups who are currently less likely to bike (including women and travelers on family/companion trips, Fig. S11), previous research has identified dedicated biking infrastructure and increased safety perceptions as two prominent areas (AitBihiOuali & Klingen, 2022; Graystone et al., 2022). The success of infrastructure in increasing bike mode shares has been demonstrated in cities in Europe (Kraus & Koch, 2021; Lanzendorf & Busch-Geertsema, 2014) and New York (AitBihiOuali & Klingen, 2022). Access to cargo bikes and electric bikes has been shown to increase biking mode choice for trips with small children (Bjørnarå et al., 2019; Riggs, 2016), and electric bikes can be competitive against cars and transit for longer trips (c.f. recommendation 2). Cargo and electric bikes are not affordable options for lower income households, some of whom have less access to transit due to residential location (Wagner et al., 2022). Attempts to increase the uptake of novel (and more expensive) biking practices must consider this constraint. The promise of increased access to sustainable mobility modes extends beyond individual trips. Ubiquitous availability of sustainable modes can reduce the necessity of car ownership for currently car dependent populations (breaking, for instance, the tight connection between household size and car ownership (Fig. S12)), and thereby facilitate larger reductions in car mode choice and mode share.

Our analysis highlights the importance of sustainable mobility in small and mid-sized cities, which are less studied than large cities (Ao & Næss, 2023). Over half of Europe's urban population lives in cities with populations of less than 1 million (Clark et al., 2018), and some of these (including Toulouse, from the cities assessed here) are predicted to grow rapidly in the next decades (Fontana, 2023). Small and medium-size cities have higher car ownership and car mode share than large cities (Fig. 2, Fig. S2), and require larger and swifter changes to reduce car dependence and use and transition to more sustainable mobility. Financial and planning resources are often more constrained in smaller cities compared to large cities. Policymakers at EU, national, and city level all have responsibility to actively facilitate sustainable mobility transitions in small and medium, as well as large cities.
Urban mobility interacts with multiple dimensions of sustainability. The way that people travel in cities has important implications for GHG emissions, air pollution, physical and mental health, and demand for energy and materials. The results of this study highlight relationships between urban form and sustainability relevant outcomes which are broadly in agreement with findings from existing literature. Key advances of the current study include illustration of the shape, strength, and spatial variation of relationships between urban form and sustainable mobility outcomes within and between different types of cities, enabling identification of commonalities and differences across countries and city types. Residential proximity to the city center is confirmed as the urban form feature with greatest potential for enabling sustainable urban mobility through its direct and indirect influence on trip distances, car ownership and urban form. Local contexts vary however, and establishing solutions appropriate to each context is crucial.

**Material and methods**

**Data**

We gather and harmonize urban mobility surveys describing trip diaries, household characteristics, and personal characteristics from nineteen cities in four European countries (Germany, France, Spain, Austria), and combine them with measures of local urban form calculated using various sources and methods. Mobility data for German cities are from the SrV 2018 survey administered by Technical University of Dresden (Hubrich et al., 2019). French mobility surveys are carried out by local authorities with technical assistance from CEREMA – the *Center for Studies and Expertise on Risks, the Environment, Mobility and Development*, which results in some standardization across surveys. The surveys are accessed from the Réseau Quetelet platform (ADISP, 2022). Mobility data from Vienna is extracted from the ‘Österreich unterwegs 2013/2014’ survey (Tomschy, R. et al., 2016), and mobility data for Madrid comes from the 2018 *Encuesta de Movilidad de la Comunidad de Madrid* (EDM2018) (Consorcio Regional de Transportes de Madrid, 2019). Further access information for all surveys is provided in the data availability statement. For modeling trip distance, we restricted the data to trips longer than 50m and shorter than 50km.

Many common variables are gathered in each survey, although some important variables, such as income are only available in certain cities. In Table 1 we summarize the variables that are gathered and their availability for each city. Table S2 in the supplementary information (SI) summarizes the spatial units used for each city, and their size distribution. For French cities and Madrid, spatial units were available at two resolutions. To make the spatial resolution closer to consistency with German cities, and because the spatial resolution was in some cases too small to usefully calculate urban form metrics, for these cities we used a combination of low- and high-resolution units for calculating urban form features. A threshold cut-off was used to divide larger spatial units into the smaller units, i.e. all lower resolution spatial units larger than 10km$^2$ were disaggregated into their constituent smaller units, if available. Of these smaller units, those which were matched to the same IRIS unit for population calculations were merged into one unit. This additional step reduced the number of very small spatial units with low populations. Further details of the procedure for combining low and high-resolution spatial units are
given in the SI and Figure S1. Spatial extents of cities were defined on a case-by-case basis, aiming for inclusion of an urban core and outer peripheries within commuting distances. The scope and boundary for inclusion for each city is described in further detail in the SI.

Population density for German postcodes is downloaded from (esri Deutschland, 2023) and refers to populations in 2022. Although this differs from the survey year (2018), the data are calculated based on 2011 Census data at high-resolution and scaled using more recent lower resolution populations, and so the variation between postcodes represents the most recent (2011 Census) data at this resolution. Population density for French cities is derived from 2012 and 2017 population counts by IRIS spatial units (Insee, 2015, 2020), aggregated or matched via spatial join as appropriate to the custom spatial units used for each French city. Population density for Madrid is from data provided at census section level for 2018 (Instituto Nacional de Estadística, 2018), again aggregated or matched via spatial join as appropriate to the custom spatial units defined for Madrid. Population density data for Wien is extracted from the most recent available data at the municipality (gemeinden) level, from the 2011 Census (Statistik Austria, 2018). Built-up density is estimated by calculating the total building volume (building footprint x height) in each spatial unit using data from the EUBUCCO database (Milojevic-Dupont, N. and Wagner, F. et al., 2023), and dividing by that unit’s area. Locations of city subcenters are estimated as locations of high built-up density, based on a double threshold approach adapted from Taubenböck et al. (Taubenböck et al., 2017), where subcenters have high density both compared to the whole city, and within other locations of similar distance from the city center. Further description of the identification algorithm is given in SI. City centers are defined manually for each city, and distance from spatial units to city center and subcenters is calculated for each spatial unit using the building-volume-weighted centroid of each unit.

Land-use shares are calculated from Urban Atlas classifications of urban land uses in 2012 and 2018 (Copernicus Land Monitoring Service, 2016, 2020), with the year closest to the survey date for each city used. Aggregated categories of ‘urban’, ‘urban fabric’, ‘commercial’, and ‘road’ are classified based on the original classes defined (European Commission, 2020). Urban fabric is an aggregation of the continuous and discontinuous urban fabric classes. ‘Commercial’ refers to ‘Industrial, commercial, public, military and private units’, ‘Construction sites’, ‘Airports’ and ‘Port areas’. ‘Urban’ contains all urban fabric and commercial uses, and also leisure facilities, roads, and urban green areas. The urban areas are used to calculate adjusted ‘urban population density’ and ‘urban built-up density’, by dividing the raw densities by the urban land-use share for each spatial unit. This provides a better estimate of local densities, especially for large spatial units which contain concentrated human settlements and large non-urban areas. Three street design metrics are calculated using the ‘stats’ module of the OSMnx python package (Boeing, 2017) - Street Intersection Density, Average Street Length, and Streets per node.
### Table 1
Household, personal, trip, and urban form features collected for each city

<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential location</td>
<td>Local administrative unit.</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>Further details on definition in text</td>
<td></td>
</tr>
<tr>
<td>Household Size</td>
<td>Number of members per household</td>
<td>All</td>
</tr>
<tr>
<td>Household Monthly Income</td>
<td>In income bands which vary across surveys</td>
<td>DE, Clermont, Toulouse, Paris</td>
</tr>
<tr>
<td>Household Vehicle Availability</td>
<td>For cars, bicycles, and motorbikes/mopeds</td>
<td>All</td>
</tr>
<tr>
<td>Household Weight</td>
<td>Weighting factor by household</td>
<td></td>
</tr>
<tr>
<td><strong>Personal Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Numeric</td>
<td>All</td>
</tr>
<tr>
<td>Sex</td>
<td>Male; Female</td>
<td>All</td>
</tr>
<tr>
<td>Mobility Constraints</td>
<td>Dummy variable</td>
<td>DE, ES, Paris</td>
</tr>
<tr>
<td>Occupation</td>
<td>Categorical, harmonized</td>
<td>All</td>
</tr>
<tr>
<td>Education</td>
<td>Categorical, categories vary by survey</td>
<td>All</td>
</tr>
<tr>
<td>Driving License</td>
<td>Dummy variable</td>
<td>All</td>
</tr>
<tr>
<td>Transit Subscription</td>
<td>Dummy variable</td>
<td>All</td>
</tr>
<tr>
<td>Car Parking at Work</td>
<td>Dummy variable</td>
<td>FR, AT</td>
</tr>
<tr>
<td>Person Weight</td>
<td>Weighting factor by person</td>
<td>All</td>
</tr>
<tr>
<td><strong>Trip Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>Reported or measured trip distance in km</td>
<td>All</td>
</tr>
</tbody>
</table>

DE = Germany, AT = Austria (only one city, Wien), ES = Spain (only one city, Madrid), FR = France
<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode Choice</td>
<td>Categorical, aggregated to five modes: walk, bike, car, public transit, motorbike (2-wheel)</td>
<td>All</td>
</tr>
<tr>
<td>Origin Purpose</td>
<td>Categorical, aggregated to: Home, Work, School, Companion, Shopping, Personal, Other</td>
<td>All</td>
</tr>
<tr>
<td>Destination Purpose</td>
<td>Same as Origin Purpose</td>
<td>All</td>
</tr>
<tr>
<td>Duration</td>
<td>Reported in minutes</td>
<td>All</td>
</tr>
<tr>
<td>Origin location</td>
<td>Local administrative unit</td>
<td>All</td>
</tr>
<tr>
<td>Destination location</td>
<td>Local administrative unit</td>
<td>All</td>
</tr>
<tr>
<td>Day of Week</td>
<td>Weekend days excluded</td>
<td>All</td>
</tr>
<tr>
<td>Season</td>
<td>Defined by month, Spring starting March 1, etc.</td>
<td>All</td>
</tr>
<tr>
<td>Accompanying householders</td>
<td>Number of other persons present for trip</td>
<td>DE</td>
</tr>
<tr>
<td>Trip Weight</td>
<td>Weighting factor by trip</td>
<td>DE, AT, ES</td>
</tr>
<tr>
<td>Urban Form Features</td>
<td><em>Calculated per local administrative unit,</em> unless otherwise specified</td>
<td>All</td>
</tr>
<tr>
<td>Density</td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>Population Density</td>
<td>Residents / km²</td>
<td>All</td>
</tr>
<tr>
<td>Built-up Density</td>
<td>Building volume (m3) / km²</td>
<td>All</td>
</tr>
<tr>
<td>Destination accessibility</td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>Distance to City Centre</td>
<td>City centers manually defined for each city</td>
<td>All</td>
</tr>
<tr>
<td>Distance to Closest Subcenter</td>
<td>City subcenters identified using method of Taubenböck et al (2017)</td>
<td>All</td>
</tr>
<tr>
<td>Time to Transit</td>
<td>Mean walking time in mins to closest stop</td>
<td>DE</td>
</tr>
</tbody>
</table>

DE = Germany, AT = Austria (only one city, Wien), ES = Spain (only one city, Madrid), FR = France
<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land-use Mix</td>
<td>Summarized to % commercial, % urban fabric, % urban, using data from Urban Atlas 2018 (Copernicus Land Monitoring Service, 2020)</td>
<td>All</td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td><em>All metrics calculated on the basis of the driving street network within 1.25km buffer of weighted centroid of each administrative unit</em></td>
<td>All</td>
</tr>
<tr>
<td>Street Intersection Density</td>
<td>Number of street intersections / km²</td>
<td>All</td>
</tr>
<tr>
<td>Average Street Length</td>
<td>Mean length of all streets</td>
<td>All</td>
</tr>
<tr>
<td>Streets per node</td>
<td>Mean count of streets per node</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>DE = Germany, AT = Austria (only one city, Wien), ES = Spain (only one city, Madrid), FR = France</td>
<td></td>
</tr>
</tbody>
</table>

**Methods**

We use gradient boosting decision tree (GBDT) regression and classification models from the *XGBoost* python library (Chen & Guestrin, 2016) to predict the average trip distance per residential postcode, car ownership per households, and mode choice per individual trip. GBDT models are well suited for this analysis as they can detect non-linear dependencies, which previous studies have shown to exist between urban form and mobility metrics (Ding et al., 2019; Wagner et al., 2022), they have shown a strong robustness against multicollinearity of input features (Ding et al., 2018), and they can identify local variations in effects of urban form (Wagner et al., 2022). The contribution of explanatory variables (features) to model predictions for each observation are extracted using the Shapley additive explanation (SHAP) values (Lundberg et al., 2020). Overall feature importance (Fig. 3) is calculated by summing the absolute SHAP values for each feature and observation, while local explainability (i.e. the contribution of a feature to model prediction over all feature values) scatter plots visualize the relationship between feature values and predicted outcomes. A literature review of ML models used for analyzing urban mobility patterns and emissions is provided in the SI.

One advantage of SHAP values for explaining model results is that we have an importance value that corresponds to every feature value, and these can be combined and visualized in order to demonstrate how the influence of a feature on the model outcome varies across the distribution of values of that feature. For models that have a spatial dimension, we can also use SHAP values to illustrate how feature influences (such as urban form features) vary spatially (as in Fig. 4 in the main text).
The identified relationships do not have a causal basis, so interpretation of SHAP values must be cautious in assessing whether an underlying causal mechanism may be present. One barrier to distinguishing between causal relationships and spurious correlations in the data is residential self-selection bias, where for example households who prefer not to own a car choose to locate in urban areas where meeting daily needs is less car dependent. In such cases, it is not (only) the urban form which encourages more households to live car-free, but also an inherent preference for avoiding car ownership (Stevens, 2017).

Model accuracy scores are reported for all models in SI Table S5, including accuracy comparisons with multivariate (trip distance) and logistic (car ownership and mode choice) regression models. To generate the accuracy scores, we used repeated k-fold cross validation for the trip distance model with 5 folds and 10 repeats, k-fold cross validation for car ownership with 9 folds, and to mitigate against hierarchical data leakage (Hillel et al., 2021) for the models of mode choice we used grouped k-fold cross validation with 9 folds, grouping by person, so that trips by the same person would never be present in both training and testing data. Feature importance (SHAP) values are estimated as the mean across all repeats. This improves the robustness of the feature importance estimates and reduces the influences of random effects associated with the stochastic GBDT models. To identify hyper-parameters which gave the best model accuracy, we used a grid search method, with variation of number of estimators, maximum tree depth, and learning rate.

**Declarations**

**Acknowledgements**

P.B. acknowledges funding from the European Commission grant no. 101027476.

**Data availability statement**

Survey data for German cities can be accessed through contact with the ‘SrV’ survey managers at TU Dresden, and the appropriate transit agencies in each city. For French cities, the survey data can be requested from the Réseau Quetelet data hosting platform. Survey data for Wien and the rest of Austria can be requested from the Federal Ministry using instructions at https://www.bmk.gv.at/themen/verkehrsplanung/statistik/oesterreich_unterwegs/methodendaten.html. Survey data for Madrid are openly available at https://datos.comunidad.madrid/catalogo/dataset/resultados-edm2018. Urban form data collected for this study will be made freely available in the ‘outputs’ directory of the project repository at https://github.com/peterberr/sufficcs_mobility.

**Code availability statement**

Code developed for this study will be made freely available upon in the project repository at https://github.com/peterberr/sufficcs_mobility.
Author contribution statement

P.B. and F.C. conceptualized the research. P.B., F.N., F.W. designed the method. P.B prepared the data and performed the analysis. P.B. drafted the manuscript with contributions from all co-authors.

References


50. Liotta, C., Viguié, V., & Creutzig, F. (2022). Policy portfolios can reduce GHG emissions in urban transport in 120 cities by 20% while improving welfare. Research Square. https://doi.org/10.21203/rs.3.rs-2131432/v1


Figures

Figure 1

*Summary of mode share by travel distance and population density in each city. Cities are ordered on the x-axis by population density, and the density in each city is indicated by the line graph and the second y-axis.*
Figure 2

a) Car ownership and b) car mode share vs household monthly income for cities with income data. ‘DE, mid’ refers to mid-sized German cities (Frankfurt, Dusseldorf, Leipzig, and Dresden). ‘DE, small’ refers to small German cities (Kassel and Magdeburg). Potsdam is a small German city, but is plotted separately as the data for Potsdam differ considerably from Kassel and Magdeburg
Figure 3

Ranking of feature importance from models of a) average trip distance, b) car ownership, and c) trip mode choice. Urban form features are defined at the residential location for a) and b), and at the point of departure for c). Feature importance is calculated as absolute SHAP values for individual features divided by the sum of absolute SHAP values.
**Figure 4**

_influence (SHAP values) of urban form features at residential location on average trip distance (in meters) in four capital cities. Areas without data are marked in white, and correspond to areas where no survey respondents live._

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)