
The built environment and induced transport CO₂ emissions: A double machine learning approach to account for residential self-selection

Florian Nachtigall

Technical University of Berlin
Berlin, Germany
nachtigall@tu-berlin.de

Felix Wagner

Technical University of Berlin
Berlin, Germany

Peter Berrill

Technical University of Berlin
Berlin, Germany

Felix Creutzig

Technical University of Berlin
Berlin, Germany

Abstract

Understanding why travel behavior differs between residents of urban centers and suburbs is key to sustainable urban planning. Especially in light of rapid urban growth, identifying housing locations that minimize travel demand and induced CO₂ emissions is crucial to mitigate climate change. While the built environment plays an important role, the precise impact on travel behavior is obfuscated by residential self-selection. To address this issue, we propose a double machine learning approach to obtain unbiased, spatially-explicit estimates of the effect of the built environment on travel-related CO₂ emissions for each neighborhood by controlling for residential self-selection. We examine how socio-demographics and travel-related attitudes moderate the effect and how it decomposes across the 5Ds of the built environment. Based on a case study for Berlin and the travel diaries of 32,000 residents, we find that the built environment causes household travel-related CO₂ emissions to differ by a factor of almost two between central and suburban neighborhoods in Berlin. To highlight the practical importance for urban climate mitigation, we evaluate current plans for 64,000 new residential units in terms of total induced transport CO₂ emissions. Our findings underscore the significance of spatially differentiated compact development to decarbonize the transport sector.

1 Introduction

The built environment plays a critical role in facilitating or hindering the decarbonization of urban transport [1]. The location and compactness of new residential development shape and constrain residents' future travel behavior. As such, they limit the mitigation potential of mobility-related lifestyle changes and potentially lock in transport-related emissions for decades [2]. Urban sprawl and suburban development, in particular, have been widely criticized for increasing car dependence and travel demand, hindering the decarbonization of urban transport [3]. Many studies have observed that residents of urban (central, higher-density, mixed-use) neighborhoods tend to drive less and instead walk [4, 5, 6, 7], bike [4], and use public transit more [8, 9] than residents of suburban (non-central, lower-density, single-use residential) neighborhoods. Yet, it is ambiguous to what extent these differences can be attributed to the built environment itself, as opposed to pre-existing differences in travel preferences between residents of urban and suburban neighborhoods. These differences are a result of a process known as residential self-selection [10], in which people choose their place of residence based on, among other things, locally available transport options matching their pre-existing travel preferences. Failing to account for residential self-selection and the resulting differences in travel preferences among residents of different neighborhoods can lead to falsely

attributing observed differences in travel behavior to the built environment alone, overestimating its impact and drawing biased conclusions [11].

Two strands in literature aim to uncover the built environment’s influence on travel behavior, one focusing on the nonlinearity of the relationship using machine learning methods, in particular decision tree ensembles [12, 13], the other focusing on disentangling the influence from the confounding effect of residential self-selection using statistical methods such as statistical control, propensity score matching, and sample selection [14]. We aim to bring both approaches together.

We explore the potential of the causal inference method *double machine learning* (DML) [15] to examine the non-linear effect of the built environment on travel behavior from mobility survey data while accounting for confounding factors. We assess the built environment’s impact across multiple, continuous dimensions, specifically the 5Ds [16], density, diversity, design, destination accessibility, and distance from transit, which all are presumed to have independent effects on travel behavior [11]. We control for the non-linear influence of socio-demographics and travel-related attitudes that are responsible for residential self-selection [14]. We examine the impact on travel behavior in terms of travel-related emissions to account for changes in both, travel distances and mode choice, allowing for a comprehensive assessment of the climate-related impacts of the built environment. We demonstrate our approach using the city of Berlin, Germany, as a case study and discuss local implications for low-carbon urban planning. Our research questions are the following:

1. What is the isolated effect of the built environment on travel behavior and induced CO₂ emissions when accounting for residential self-selection for each neighborhood in Berlin?
2. How does the effect decompose into the 5Ds and how is it moderated by socio-demographics and travel-related attitudes?
3. What are the induced transport CO₂ emissions of currently planned housing projects in Berlin and how can the spatial allocation be improved to most effectively reduce emissions?

2 Methods

2.1 Preprocessing

Data sources. Representative travel behavior per neighborhood is obtained from travel diaries of the German national mobility survey "Mobilität in Städten – SrV 2018" [17], which includes more than 100,000 trips for Berlin. To describe the built environment, we use only publicly available data from OpenStreetMap [18] on the street network and points of interest, from the public transit operator VBB on transit accessibility (GTFS data) [19] and from Berlin’s open data portal [20] on population density, land-use, street greenery, street space allocation, and election results. See figure 4 for a visualization of the built environment of Berlin. Lastly, to convert mode-specific travel kilometers to emissions, we use emissions factors from the International Transport Forum (ITF) [21] (see table 2).

Unit of analysis. We approximate the representative average travel behavior per neighborhood by averaging the surveyed and preprocessed travel behavior of all households. We assume that neighborhoods, in our case zip code areas, are sufficiently homogeneous in terms of their built environment to be able to detect a consistent impact on the travel behavior of its residents. This approach has the benefit of reducing noise related to sampling from a wide range of individual daily travel patterns and only examining to what degree the average travel behavior differs between neighborhoods.

2.2 Causal inference

Confounding effects. Travel preferences confound the influence of the built environment on travel behavior because of residential self-selection. The relationship is visualized as a directed acyclic graph (DAG) in figure 1. The confounding effect can be accounted for by controlling for socio-demographic traits and travel-related attitudes [14]. Here, we use information on age, income, education, and household size from the mobility survey to describe the socio-demographic composition of neighborhoods. Because travel-related attitudes are not directly captured in the survey, we use information on ownership of transport means as proxies, specifically car, bike, driving license, and transit subscription ownership. While this is not a comprehensive characterization of travel-related attitudes, it has the advantage of capturing the temporal lag of travel-related attitudes through vehicle ownership, particularly car ownership [22], that not only locks in travel behavior but also influences

residential choice [23, 24]. Complementary to transport ownership, we capture general attitudes towards sustainable mobility and other environmental issues according to recent election results where the provision of sustainable transport infrastructure was in the public focus. Refer to table 3 for a brief description of all attributes considered with respect to travel preferences.

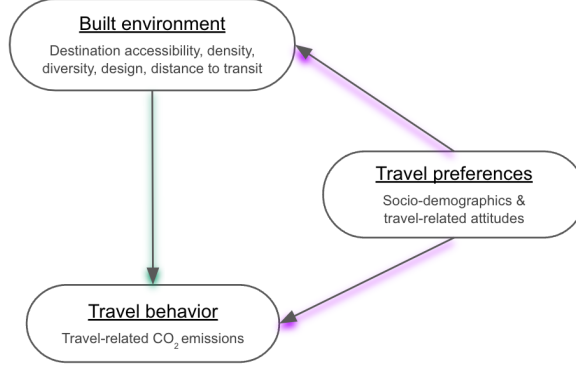


Figure 1: Directed acyclic graph (DAG). Travel preferences confound the effect of the built environment on travel behavior via residential self-selection. Operationalization of concepts as specified in the grey text.

Treatment encoding. To examine how the built environment of a specific neighborhood shapes travel behavior, we first determine how the local built environment is different from the city average. We characterize the difference along the 5Ds, density, diversity, design, destination accessibility, and distance from transit, for each of which we engineer at least one feature, guided by previous studies (see table 4). To quantify how this difference impacts travel behavior, we define the treatment level as the difference between the neighborhood built environment and the city average built environment.

Model selection. Given continuous treatment along multiple treatment dimensions, we choose a CausalForest-based DML estimator to estimate the causal non-linear effect of the built environment on travel behavior. DML [15] consists of two stages: First, to account for the confounding effect, the outcome and treatment are being predicted from the controls using any appropriate machine learning model. Then, the treatment residuals are used to fit the outcome residuals yielding a debiased treatment effect estimate. The controls are also included to account for any moderating effects in explaining the heterogeneity of the treatment effect. We use the DML open source implementation of the EconML package [25]. For the first stage, we use XGBoost [26] to model the non-parametric relationship between travel preferences and travel behavior and the built environment. We choose hyperparameters based on a random search using 5-fold cross validation, resulting in 1000 tree estimators, a tree depth of 6, and a learning rate of 0.01. For the final model, we use a CausalForest [27] with 100 tree estimators.

Effect heterogeneity & composition. To analyze the heterogeneity of the built environment’s effect on travel behavior, we calculate SHAP values [28] for the constant marginal effect estimation of the final stage model. This allows us to examine the moderating influence of confounding socio-demographic traits and travel-related attitudes on the effect. We are interested in the subgroups of households for which the built environment may have a particularly large or small effect. To avoid reporting of spurious correlations that the model picked up, we repeat the effect estimation 10 times, for all 5D dimensions combined and each individually, and only report moderators with a clear and consistent influence across all iterations. Further, we decompose the total effect of the built environment into the 5D dimensions and discuss the main drivers for low-carbon urban development. To make the 5D dimensions comparable in terms of their contribution to the overall marginal effect, we standardize all built environment characteristics and ensure consistent direction of feature values.

2.3 Urban planning case study.

To highlight the practical importance for urban climate mitigation, we evaluate current plans for 64,000 new residential units in Berlin in terms of total induced transport emissions. We compare the induced emissions of three alternative densification scenarios: (1) transit-oriented development

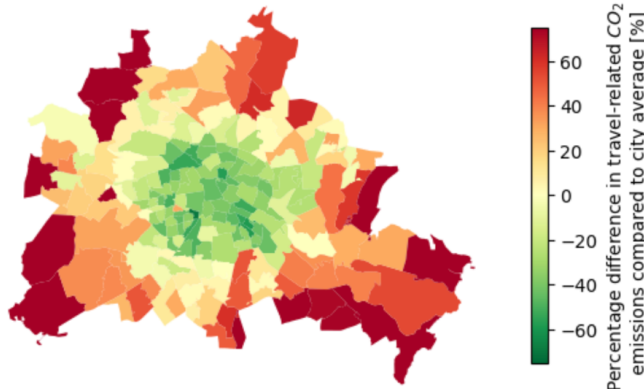
focused on neighborhoods whose residents have an average walk time of less than 7 minutes to the nearest rail station, (2) transit-oriented development focused on centrally located neighborhoods, specifically neighborhoods connected by or located within the commuter rail line that circles central Berlin, the so-called “Ringbahn”, and (3) densification of low-emission neighborhoods where the built environment has the largest reducing impact on emissions according to our estimations. We assume that the housing units will be evenly distributed among the targeted neighborhoods for all three scenarios. We characterize the potential emissions savings from each strategy compared to the average household emissions. We assume that the built environment remains unchanged, i.e., we do not consider potential changes related to the new residential developments.

3 Results

Isolated effect. The built environment has a considerable effect on travel behavior and related emissions after controlling for residential self-selection. In the center of Berlin, the built environment facilitates a 40% decrease and in the outskirts a 50% increase in travel-related emissions compared to the city average (see figure 2). Overall, half of the observed differences in neighborhood emissions are due to the built environment according to our CausalForest-based DML estimator. The remaining residuals may partly be due to residential self-selection and an insufficient characterization of the built environment. Overall our features explain the variation of neighborhood emissions with a coefficient of determination (R^2) of 0.85, suggesting a good characterization of travel preferences and the built environment. The nuisance scores of the first stage DML models indicate a high degree of confounding, with the R^2 being 0.76 and 0.56 for fitting the target and treatment, respectively.

Effect heterogeneity. The estimated effect of the built environment is moderated by socio-demographics and travel-related attitudes. The household size, income, age and car ownership are positively associated with the impact of distance to the city center on travel-related emissions meaning that the impact of the built environment may be larger for households that are relatively large, old, high-income, or highly car-owning (see figure 5). On the other hand, environmentally friendly attitudes are associated with a lower effect of the built environment. The other confounding factors do not exhibit a consistent impact on the heterogeneity across multiple iterations with different seeds and are thus excluded from our results.

Effect composition. From the 5Ds, destination accessibility has the largest effect on travel-related emissions. 73.7% of the built environment’s influence is determined by the distance to the center, to the subcenter and the local neighborhood job accessibility (see table 1). The second most important D is density, with the adjusted population density being responsible for 15.2% of the total effect. While design and distance from transit have a small effect, 6.4% and 4.3% respectively, diversity in terms of the mixed-use land share has no meaningful impact on travel-related emissions according to our approach.



5D	Feature name	Effect share
Destination accessibility	Distance to center	51.2%
	Distance to subcenter	15.2%
	POI density index	11.1%
Density	Population density	11.4%
Diversity	Land use	0.3%
Design	Car-friendliness index	-
	Walkability index	6.4%
Distance to transit	Transit accessibility index	4.3%

Figure 2: Spatial heterogeneity of the built environment’s effect on travel-related CO₂ emissions.

Table 1: Decomposition of the built environment’s overall constant marginal treatment effect.

Urban planning case study. Due to location-dependent built environment effects, the planned settlements in Berlin are expected to increase average household travel-related emissions by 16.8%

above the city average (see figure 3). Alternative residential development strategies that focus on densifying the center and transit-oriented development may lead to significantly lower emissions. We estimate that emissions can be reduced by up to 30.9% if residential planning prioritizes the 20 neighborhoods with the most sustainable built environment according to the model. If all neighborhoods with a good transit accessibility are targeted, specifically within 7 minutes of walking to the nearest rail station, emissions of future residents are expected to be 14% below the city average.

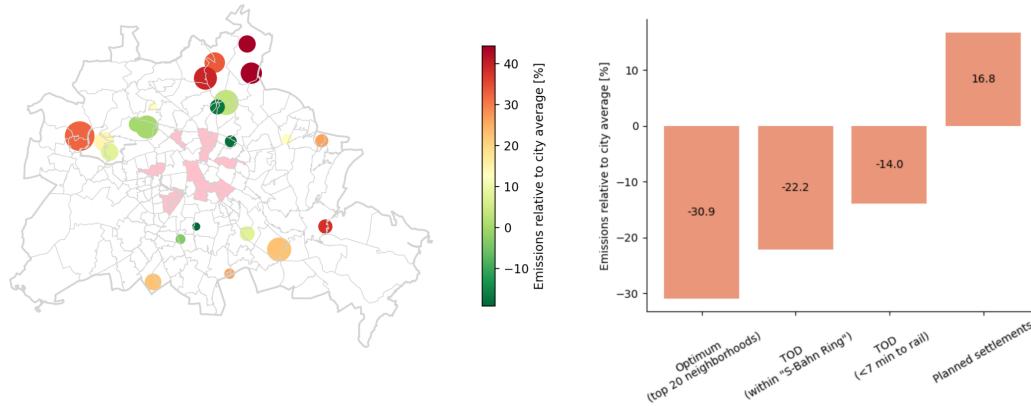


Figure 3: Induced transport CO₂ emissions of planned settlements 70% above the theoretical optimum. Left: New settlement locations are marked by circles, with color representing future residents' expected average emissions and size indicating planned housing units. Neighborhoods with the largest reducing effect on emissions, targeted by the "Optimum" policy, are marked in light pink. Right: Comparison of different residential planning strategies, including transit-oriented development (TOD) and a densification according to the model ("Optimum"), in terms of average household travel-related emissions relative to the city average.

4 Discussion & Conclusion

With the need to reduce transportation emissions and curb global boiling, growing cities face the urgent question of where to locate new residents to minimize travel demand and related emissions. Thus, having a spatially explicit estimate of the built environment's impact on transport emissions at potential new residential development planning sites is indispensable for evidence-based low-carbon urban planning.

In Berlin, the impact of the built environment on travel-related emissions is tremendous. Household emissions differ by a factor of two between the city center and the outskirts because of the built environment. This disparity is mainly due to the different accessibility of destinations (74%). We find that the effect of the built environment is largest for households that are relatively large, old, high-income, or car-owning. According to our calculations, the induced emissions of the currently planned 66,000 residential units are 70% above the optimum. Alternative compact or transit-oriented development strategies would lead to significantly lower emissions.

While we believe that the overall effect magnitude is robust, we want to emphasize that the precise estimates of induced emissions are subject to uncertainty for three main reasons. First, causal inference assumptions are partially violated, most importantly we do not account for spatial interference of treatments and spatial confounding effects [29]. Second, travel-related attitudes are likely affected to some degree by the built environment and past travel behavior and thus interdependent instead of being an exogenous predisposition [30, 31]. Third, by aggregating travel behavior and built environment characteristics for each zip code area, we mask heterogeneous distributions and nuances in the disaggregated data, potentially leading to a biased estimate (MAUP effect [32, 33]).

In conclusion, double machine learning has the potential to greatly facilitate and scale the estimation of travel demand induced by the built environment. Although the effect estimates are subject to some uncertainty, it can provide cities with a cost-effective tool to establish a starting point for evidence-based sustainable urban planning.

Acknowledgments and Disclosure of Funding

This work received funding from the CircEular project of the European Union's Horizon Europe research and innovation program under grant agreement 101056810.

References

- [1] P. Jaramillo, S. Kahn Ribeiro, P. Newman, S. Dhar, O.E. Diemuodeke, T. Kajino, D.S. Lee, S.B. Nugroho, X. Ou, A. Hammer Strømman, and J. Whitehead. Transport. In P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, editors, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, book section 10. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.
- [2] F Creutzig, P Agoston, J C Minx, J G Canadell, R M Andrew, C Le Quéré, and S Dhakal. Urban infrastructure choices structure climate solutions. *Nature Climate Change*, 6(12):1054–1056, 2016.
- [3] Frans Dieleman and Michael Wegener. Compact city and urban sprawl. *Built environment*, 30(4):308–323, 2004.
- [4] Veerle Van Holle, Benedicte Deforche, Jelle Van Cauwenberg, Liesbet Goubert, Lea Maes, Nico Van de Weghe, and Ilse De Bourdeaudhuij. Relationship between the physical environment and different domains of physical activity in european adults: a systematic review. *BMC public health*, 12(1):1–17, 2012.
- [5] Gavin R McCormack and Alan Shiell. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *International journal of behavioral nutrition and physical activity*, 8:1–11, 2011.
- [6] David W Barnett, Anthony Barnett, Andrea Nathan, Jelle Van Cauwenberg, and Ester Cerin. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. *International journal of behavioral nutrition and physical activity*, 14(1):1–24, 2017.
- [7] Ester Cerin, Andrea Nathan, Jelle Van Cauwenberg, David W Barnett, and Anthony Barnett. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. *International journal of behavioral nutrition and physical activity*, 14(1):1–23, 2017.
- [8] Laura Aston, Graham Currie, Alexa Delbosc, Md Kamruzzaman, and David Teller. Exploring built environment impacts on transit use—an updated meta-analysis. *Transport reviews*, 41(1):73–96, 2021.
- [9] Anna Ibraeva, Gonçalo Homem de Almeida Correia, Cecília Silva, and António Pais Antunes. Transit-oriented development: A review of research achievements and challenges. *Transportation Research Part A: Policy and Practice*, 132:110–130, 2020.
- [10] Xinyu (Jason) Cao, Patricia L. Mokhtarian, and Susan L. Handy. Examining the impacts of residential self-selection on travel behaviour: A focus on empirical findings. 29(3):359–395. Publisher: Routledge _eprint: <https://doi.org/10.1080/01441640802539195>.
- [11] Xiaodong Guan, Donggen Wang, and Xinyu Jason Cao. The role of residential self-selection in land use-travel research: a review of recent findings. 40(3):267–287. Publisher: Routledge _eprint: <https://doi.org/10.1080/01441647.2019.1692965>.
- [12] Mahdi Aghaabbasi and Saksith Chalermpong. Machine learning techniques for evaluating the nonlinear link between built-environment characteristics and travel behaviors: a systematic review. *Travel behaviour and society*, 33:e100640–e100640, 2023.
- [13] Jason Cao and Tao Tao. Using machine-learning models to understand nonlinear relationships between land use and travel. *Transportation Research Part D: Transport and Environment*, 123:103930, 2023.

- [14] Patricia L Mokhtarian and Xinyu Cao. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological*, 42(3):204–228, 2008.
- [15] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.
- [16] Reid Ewing and Robert Cervero. Travel and the built environment: A meta-analysis. *Journal of the American planning association*, 76(3):265–294, 2010.
- [17] Stefan Hubrich, Frank Ließke, Rico Wittwer, Sebastian Wittig, and Regine Gerike. Methodenbericht zum forschungsprojekt „mobilität in städten–srv 2018 “. 2019.
- [18] OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>, 2017.
- [19] VBB Verkehrsverbund Berlin-Brandenburg GmbH. GTFS data berlin-brandenburg (VBB).
- [20] Offene daten berlin | offene daten lesbar für mensch und maschine. das ist das ziel.
- [21] P Cazzola and P Crist. Good to go? assessing the environmental performance of new mobility. 2020.
- [22] Veronique Van Acker, Patricia L Mokhtarian, and Frank Witlox. Car availability explained by the structural relationships between lifestyles, residential location, and underlying residential and travel attitudes. *Transport Policy*, 35:88–99, 2014.
- [23] Tao Lin, Donggen Wang, and Xiaodong Guan. The built environment, travel attitude, and travel behavior: Residential self-selection or residential determination? 65:111–122.
- [24] Joachim Scheiner. Residential self-selection in travel behavior: Towards an integration into mobility biographies. *Journal of Transport and Land Use*, 7(3):15–29, 2014.
- [25] Microsoft Research. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. <https://github.com/microsoft/EconML>, 2019. Version 0.x.
- [26] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [27] Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. 2019.
- [28] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
- [29] Brian J Reich, Shu Yang, Yawen Guan, Andrew B Giffin, Matthew J Miller, and Ana Rappold. A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review*, 89(3):605–634, 2021.
- [30] Petter Næss. Residential self-selection and appropriate control variables in land use: Travel studies. 29(3):293–324. Publisher: Routledge _eprint: <https://doi.org/10.1080/01441640802710812>.
- [31] Joachim Scheiner. Transport costs seen through the lens of residential self-selection and mobility biographies. 65:126–136.
- [32] A Stewart Fotheringham and David WS Wong. The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*, 23(7):1025–1044, 1991.
- [33] Ming Zhang and Nishant Kukadia. Metrics of urban form and the modifiable areal unit problem. *Transportation Research Record*, 1902(1):71–79, 2005.
- [34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Appendix

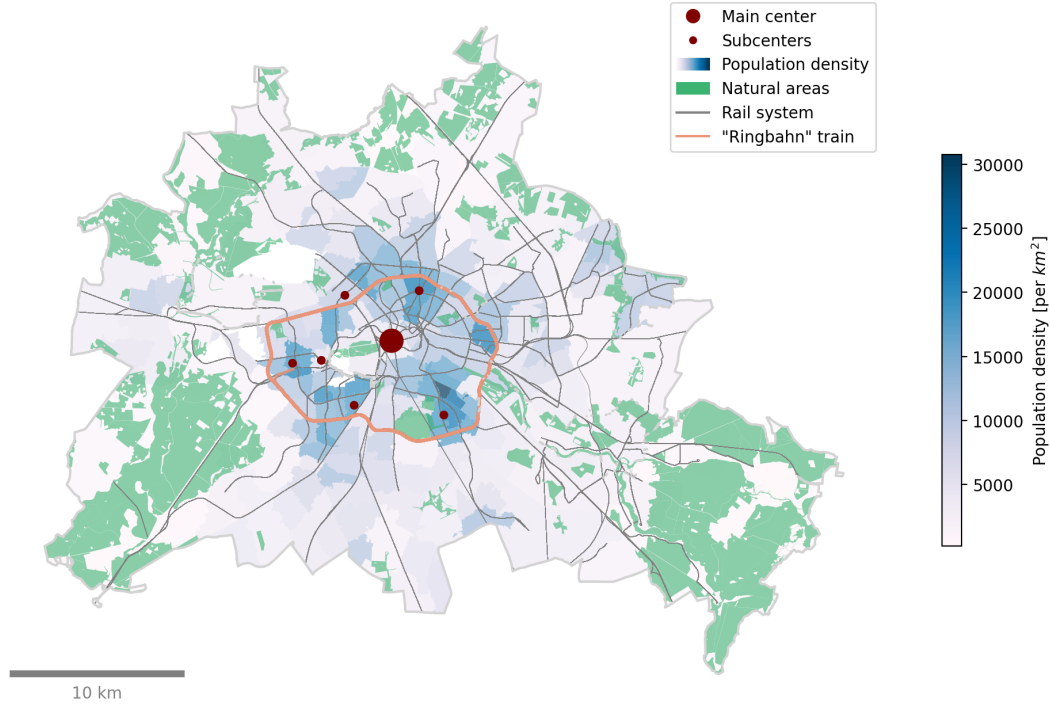


Figure 4: Built environment of Berlin, Germany. The center and subcenters, based on points of interest density, are indicated as dark red circles. Population density of neighborhoods is color coded in blue. Natural areas according to Berlin land use data [20] are marked in green. The public transportation rail network is drawn in gray, with the exception of the so-called "Ringbahn", a commuter rail line that circles central Berlin, which is highlighted in orange. We consider neighborhoods that are located outside of the "Ringbahn" and not within walking distance to be suburban.

Mode	Emissions [g CO ₂ /pkm]
Car (ICE)	162
Moped (ICE)	70
Transit	65
Bike	20
Foot	0

Table 2: Emission factors of transport modes. Central estimates of life-cycle greenhouse gas emissions of urban transport modes per person km according to the International Transport Forum (ITF) [21]. Emissions factors are expressed CO₂ equivalents and have partially been aggregated to match transport modes considered in this study (e.g. bus & metro → transit). Life-cycle emissions include a vehicle, fuel, and infrastructure component as well as operational services. ICE refers to internal combustion engine.

Category	Variable name	Description
Socio-demographics	income	Average household income
	hh_size	Average number of persons living in a household
	age	Average age of adult (>18 years) residents
	uni_share	Share of people older than 25 with university degree
Proxies for travel-related attitudes	car_ownership	Average number of private & company cars per household
	bike_ownership	Average number of bicycles owned per person
	driving_license	Average share of adults (>18 years) with driving license
	transit_subscription	Average share of people with monthly transit subscription (incl. children and people with disabilities with free ride tickets)
	green_share	Electoral share of the Green party in constituencies intersecting the neighborhood in the last regional elections

Table 3: Travel preferences. Overview of all socio-demographic traits and proxies for travel-related attitudes considered in the study including a brief description. Children are excluded from the calculation of average age, education, and driver’s license ownership to reduce the correlation between the variables and ensure that each variable describes a separate aspect.

5D’s of compact development	Feature name	Description
Destination accessibility	Distance to center	Distance to neighborhood with highest POI density
	Distance to subcenter	Least distance to any of the 10 neighborhoods with highest POI density
	POI density index	Local POI density for offices, schools, kindergarten, and universities
Density	Population density	Population density of the built-up area
Diversity	Land use	Share of mixed-use areas
Design	Car-friendliness index	Provision of expressway kilometers per capita
	Walkability index	Intersection density in the built-up area
Distance from transit	Transit accessibility index	Gravity model-based index describing the average spatio-temporal transit accessibility of a neighborhood

Table 4: Built environment characteristics. Overview of all built environment characteristics considered in the study including a brief description and mapping to the corresponding D of compact development. The street network and points of interest (POIs) are extracted from OpenStreetMap. Density measures are calculated using the built up area only. Indices are calculated using sklearn’s *StandardScaler* [34].

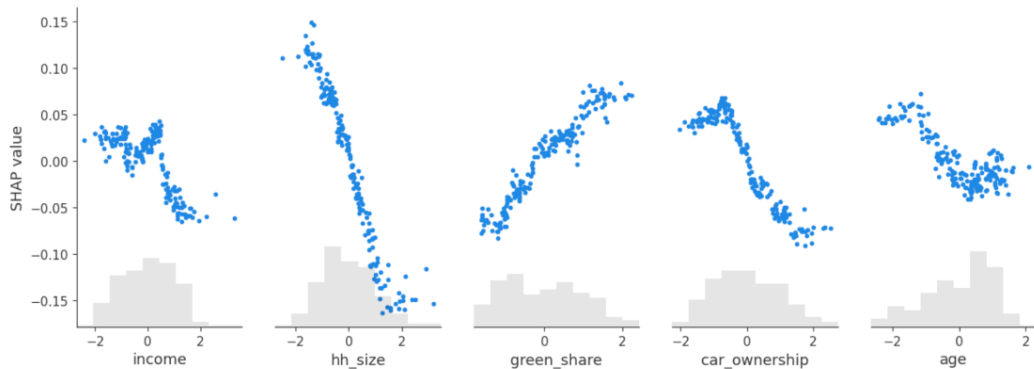


Figure 5: Household size, income, age, car ownership, and environmental attitudes consistently moderate treatment effect. SHAP dependence plot for the final stage model predicting treatment heterogeneity. Selection of confounders that have a consistent moderating effect on travel-related emissions. Confounding variables have been standardized to facilitate comparison. A SHAP value below 0 corresponds to an increased effect of the built environment for such households, whereas a SHAP value above zero corresponds to a decreased effect. Consistent zero SHAP values for all household characteristics would imply that there is no effect heterogeneity.