The built environment and induced transport $CO_2$ emissions

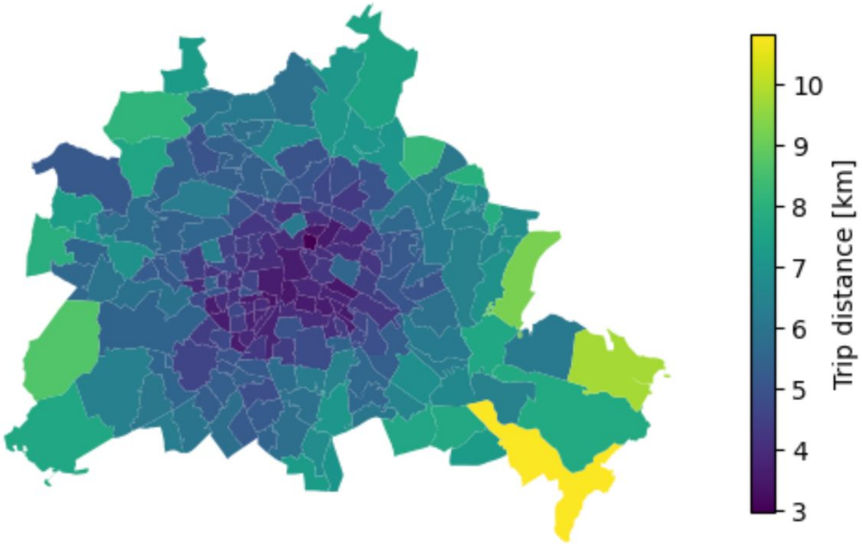
# A double machine learning approach to account for residential self-selection

Florian Nachtigall, Felix Wagner, Peter Berrill, Felix Creutzig

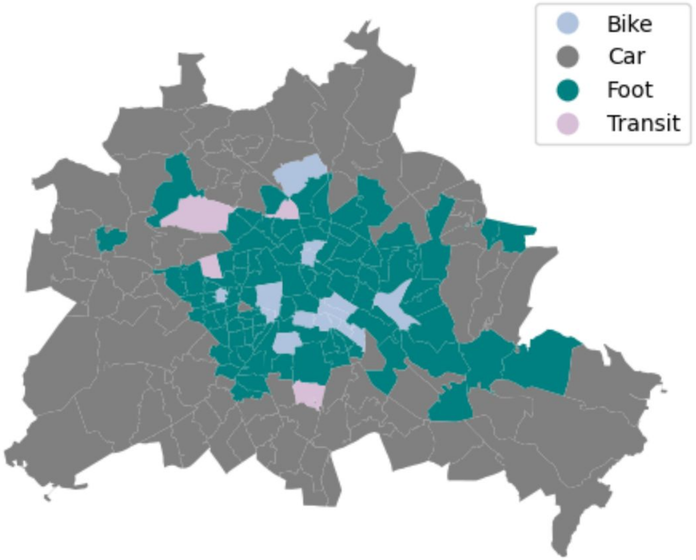

Climate relevance

# Where to locate new housing to minimize travel-related $CO_2$ emissions?

# Travel behavior differs between urban & suburban residents




Trip distance




Mode choice

Motivating question

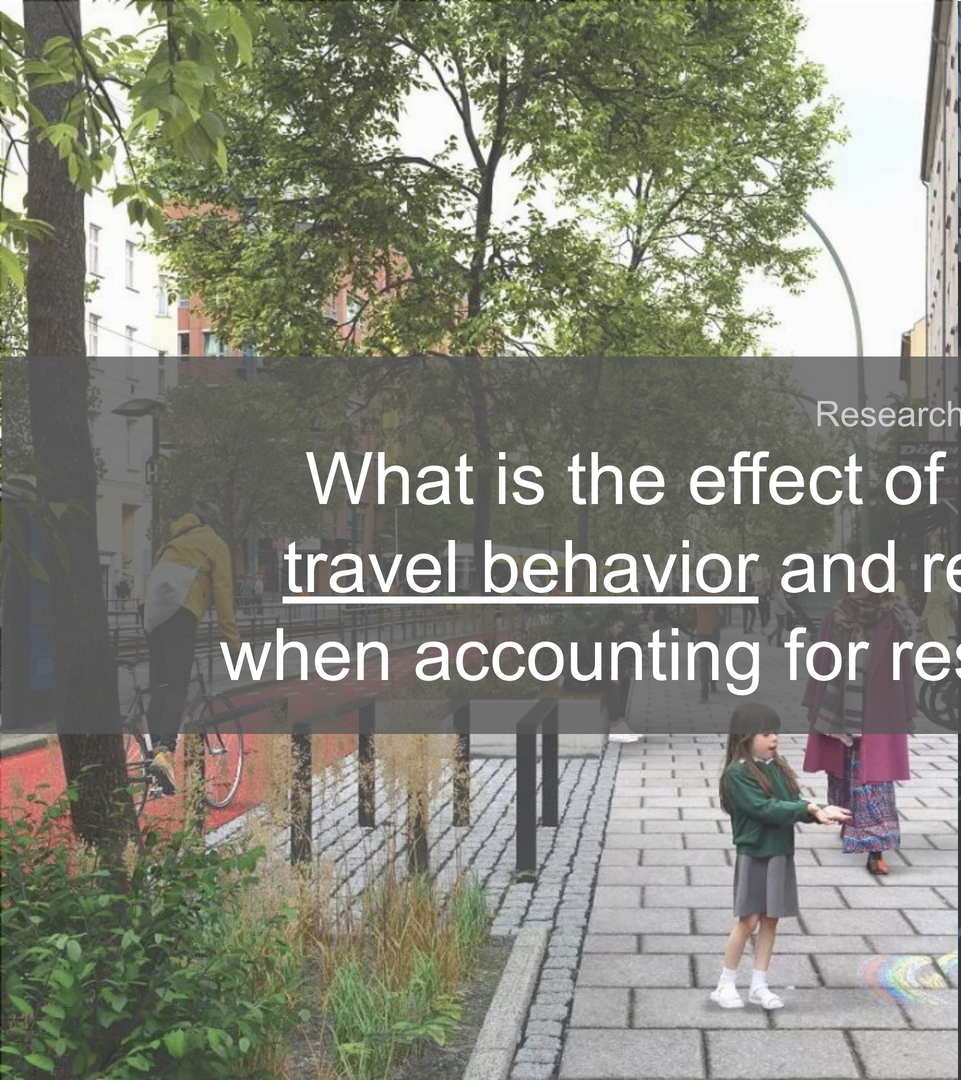
# Why do emissions differ between urban & suburban areas?

Motivating question

# Why do emissions differ between urban & suburban areas?

<u>Two possible explanations</u>

1. Different kind of people (residential self-selection)
2. Differences in the built environment



Research question

# What is the effect of <u>built environment</u> on <u>travel behavior</u> and related $CO_2$ emissions when accounting for residential <u>self-selection</u>?

# Contribution

**Existing work**

- Non-linear effect estimation wo/ confounding factors (ML methods)
- Linear (causal) effect estimation w/ confounding factors (propensity score matching & sample selection)

**Our contribution**

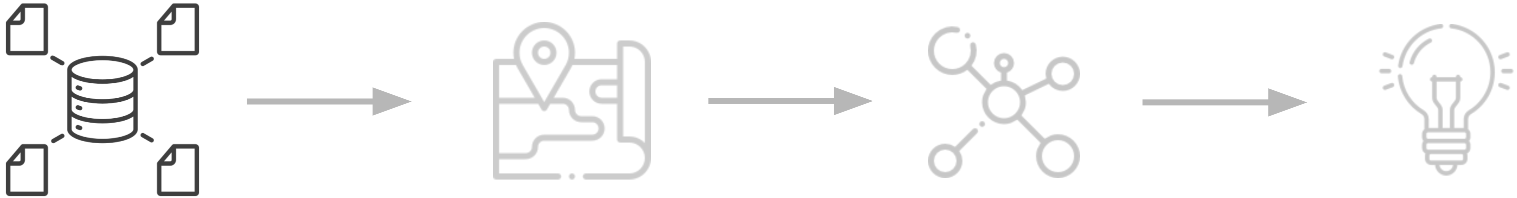
- Combining both approaches using double machine learning
    - Model nonlinearity
    - Control for confounding effects
    - Capture moderating influence and effect heterogeneity

# Methods

I. Data & preprocessing    II. Feature engineering    III. Causal inference

# Overview

**Data & preprocessing** → Feature engineering → Causal inference → Scenario modeling

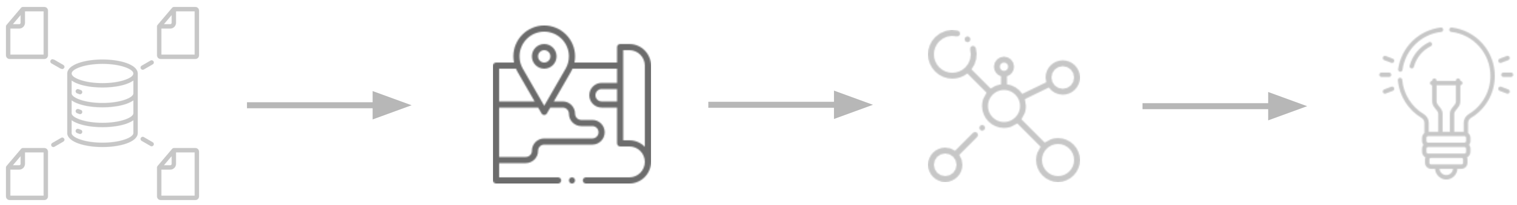


- Travel diaries from survey (32k participants)
- Calculate per household emissions based on travel distance, mode, and emission factors
- Average travel-related emissions per residential zip code

| Ori_Plz | Des_Plz | Mode | Trip_Purpose | Trip_Duration | Trip_Distance | emissions |
|---|---|---|---|---|---|---|
| 10115 | 10115 | Transit | Home-Work | 20 | 729.0 | 47.385 |
| 10115 | 10115 | Transit | Work-Home | 30 | 729.0 | 47.385 |
| 10179 | 10179 | Foot | Home-Leisure | 30 | 7000.0 | 0.000 |
| 10179 | 10179 | Foot | Leisure-Home | 10 | 254.0 | 0.000 |
| 10179 | 10179 | Foot | Leisure-Home | 30 | 7000.0 | 0.000 |
| ... | ... | ... | ... | ... | ... | ... |
| 12619 | 12169 | Car | Home-Work | 60 | 28110.0 | 4553.820 |
| 12169 | 12619 | Car | Work-Home | 60 | 26952.0 | 4366.224 |
| 12619 | 15344 | Transit | Home-Work | 70 | 22268.0 | 1447.420 |
| 12623 | 15366 | Car | Home-Shopping | 10 | 2581.0 | 418.122 |
| 15366 | 12623 | Car | Shopping-Home | 10 | 2581.0 | 418.122 |

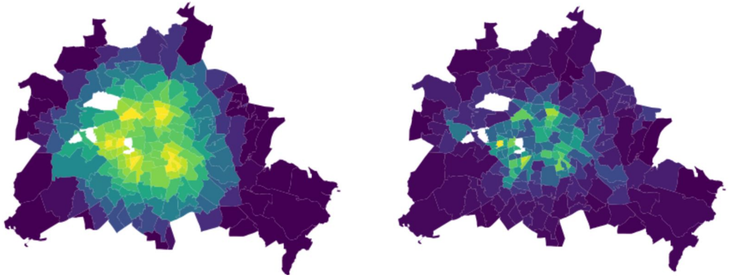
# Overview

**Data & preprocessing** → **Feature engineering** → **Causal inference** → **Scenario modeling**



| 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 mix-use areas |
| Design | Car-friendliness index | Provision of expressway kilometers per capita |
| | Walkability index | Intersection density in the built-up area |
| Distance to transit | Transit accessibility index | Gravity model-based index describing the average spatio-temporal transit accessibility of a neighborhood |

**Table 2:** ***Built environment characteristics.*** *Overview of all built environment characteristics considered in*



# Overview

**Data & preprocessing** → **Feature engineering** → **Causal inference** → **Scenario modeling**

- Select confounders (encode travel preferences)
- Define treatment as the difference to the average built environment
- DML due to multiple, continuous treatment dimensions
- XGBoost for nuisance models, CausalForest for final model (EconML implementation)
- Examine moderating effects

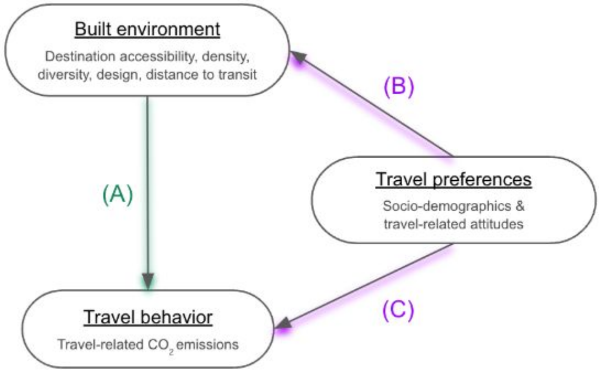
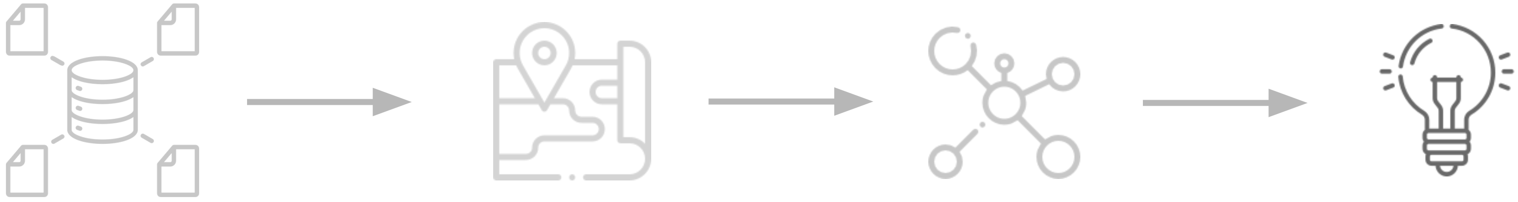



***Figure 1. Directed acyclic graph (DAG)***

*Methods*

# Overview

Data & preprocessing — Feature engineering — Causal inference — **Scenario modeling**



- Apply model to evaluate locations of planned residential projects
- Compare different urban planning strategies such as TOD

# Results

I. Causal effect estimation    II. Moderating effects    III. Scenario modeling

*Results*

# Effect of the built environment

➔ Travel-related $CO_2$ emissions differ by a factor of two between urban and suburban neighborhoods in Berlin because of the built environment

➔ Destination accessibility has the strongest impact on emissions

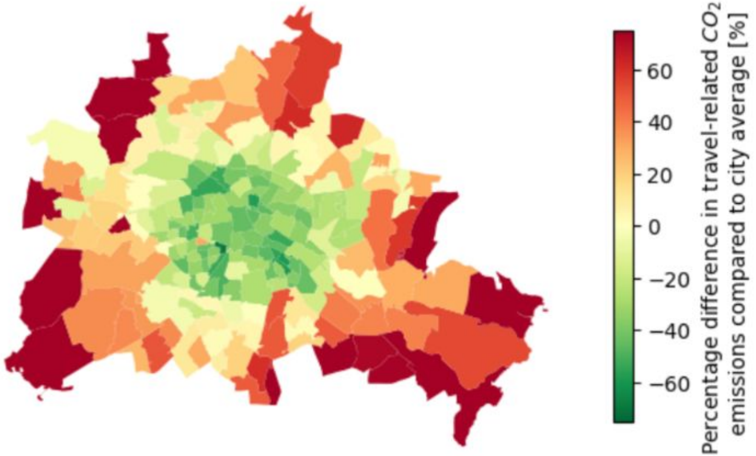



***Figure 2. Spatially-explicit effect of the built environment on travel-related $CO_2$ emissions***

| 5D's of compact development | 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% |

***Table 2. Decomposition of built environment effect.***

*Results*

# Moderating effects

- ➔ Household size, income, age, and car ownership are associated with a higher effect of the built environment
- ➔ Positive environmental attitudes with a lower effect of the built environment

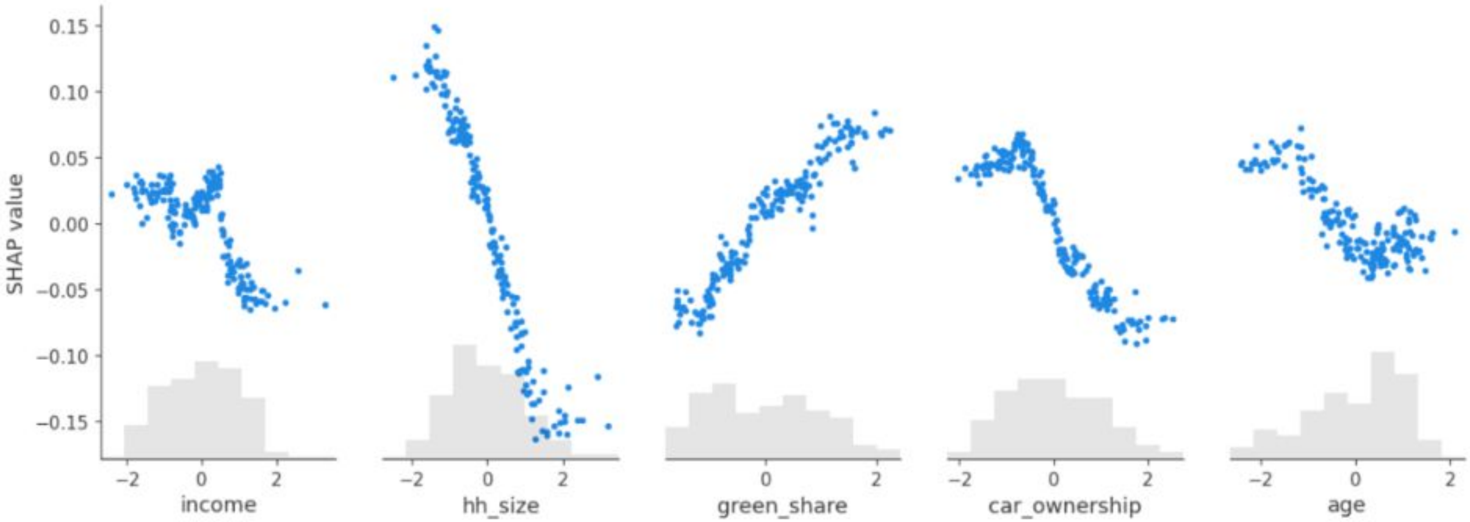



***Figure 3. Moderating influence on the built environment effect of distance to center.***

# Case study of planned residential projects

➔ Induced transport $CO_2$ emissions of planned residential projects 70% above the theoretical optimum of urban densification

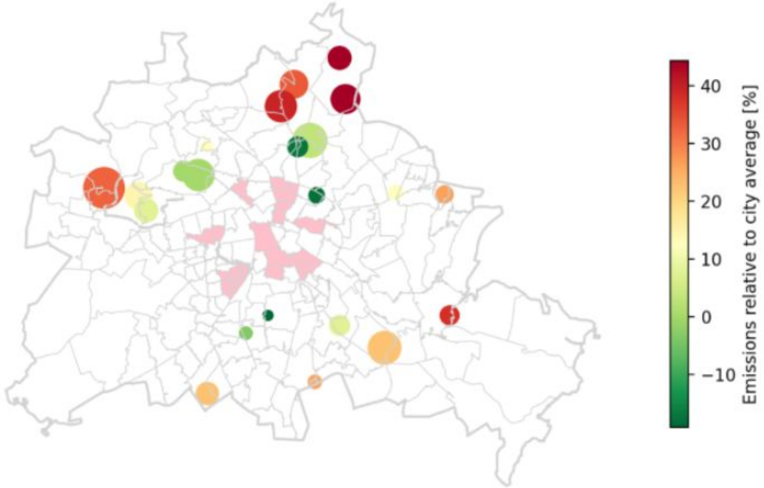



***Figure 4B. Induced transport $CO_2$ emissions of planned residential projects.***

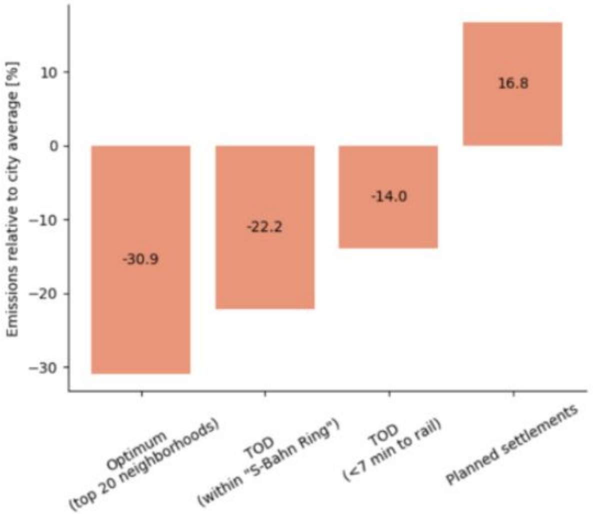



***Figure 4A. Induced transport $CO_2$ emissions of residential planning strategies.***

# Discussion & conclusion

# Advancing evidence-based low-carbon residential planning

- Large effect of the built environment on travel behavior
  - Emissions differ by a factor of two between the city center and the outskirts
  - Declining accessibility of destinations (74%) and population density (15%) drive emissions
- Moderating effects
  - Largest effect for old, high-income, and car-owning households
- Limitations masking the true effect of the built environment
  - Incomplete characterization of the built environment and travel preferences
  - Oversimplified conceptual representation (e.g. ignoring mediating effect of the built environment on travel preferences)
  - Partial violation of causal inference assumptions (e.g. spatial spillover effects)

Climate mitigation conclusion

# Compact development is key to decarbonize urban transport

Technische Universität Berlin

MCC Mercator Research Institute on Global Commons and Climate Change

The built environment and induced transport $CO_2$ emissions

# A double machine learning approach to account for residential self-selection

Florian Nachtigall, Felix Wagner, Peter Berrill, Felix Creutzig

# Thanks for listening!

https://floriannachtigall.github.io/
nachtigall@tu-berlin.de

# 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.

# Credits

- Pictures from unsplash.com
- Icons from flaticon.com

# Appendix

# Poster

## The built environment & induced transport emissions

### A double machine learning approach to account for residential self-selection

Florian Nachtigall, Felix Wagner, Peter Berrill, Felix Creutzig

NEURAL INFORMATION PROCESSING SYSTEMS

### Introduction

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 $CO_2$ 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.

**Research question**

What is the effect of built environment on travel behavior and related $CO_2$ emissions when accounting for residential self-selection?

### Methods

Use double machine learning (DML) to control for residential self-selection and obtain spatially explicit estimates of the effect of the built environment on travel-related $CO_2$ emissions for each neighborhood from observational data.

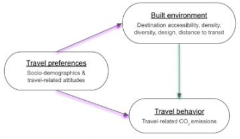



Figure 1. Directed acyclic graph (DAG) with direct (green) and confounding effect (pink).

Data & preprocessing

- Travel diaries from 2017 SrV mobility survey (32k participants in Berlin)
- Calculate emissions based on travel distance, mode, and emission factors
- Average household travel-related emissions per residential zip code

Feature engineering

- Describe built environment along "5Ds": destination accessibility, density, diversity, design, and distance to transit

Causal inference

- Estimate treatment effect of built environment from observation data
- Confounders: Account for residential self-selection using information on socio-demographics and ownership of transport means
- Treatment level: Difference to average built environment
- Model selection: DML due to multiple, continuous treatment dimensions
- XGBoost for nuisance models, CausalForest for final model (EconML implementation)

### Results

- → **Treatment effect:** Travel-related $CO_2$ emissions differ by a factor of two between urban and suburban neighborhoods in Berlin because of the built environment (see figure 2)
- → **Effect decomposition:** Declining accessibility of destinations (74%) and population density (15%) drive emissions (see table 1)
- → **Moderating effects:** Built environment effect is largest for old, high-income, and car-owning households (see figure 3)

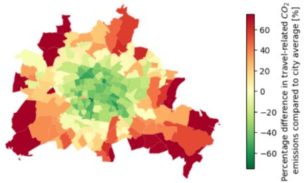



Figure 2. Spatially-explicit effect of the built environment on travel-related $CO_2$ emissions.

| 5D's of compact development | 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% |

Table 1. Decomposition of the built environment's effect.

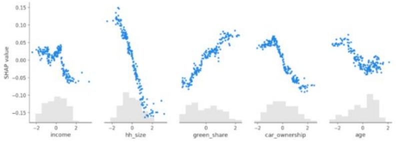



Figure 3. Moderating influence on the built environment's effect (example for distance to center).

### Case study

Assessment of planned housing projects in Berlin in terms of induced transport $CO_2$ emissions.

- → 19 of 22 location will increase emissions, on average 17% above city's current average and 70% above ideal urban densification according to model (see figure 4)

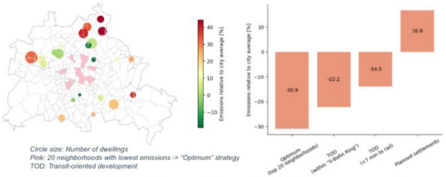



Figure 4. Induced transport $CO_2$ emissions of planned residential projects (left) and urban planning strategies (right).

### Discussion

DML can advance evidence-based low-carbon urban planning

- Spatial explicit estimates of representative travel-related $CO_2$ emissions facilitate residential planning

Compact development is key to decarbonize urban transport

- Increase destination accessibility and population density to reduce emissions
- Impact likely larger in stressed housing market as many people are not able to realize their urban preferences and use sustainable modes of transport

Limitations mask true effect of the built environment

- Oversimplified conceptual representation (e.g. ignoring mediating effect of built environment on travel preferences)
- Partial violation of causal inference assumptions (e.g. ignoring spatial spillover effects)

### Conclusion

- → Double machine learning (DML) based on mobility surveys enables spatially targeted compact development to decarbonize urban transport

See link / QR code for references, presentation recording and slides.

MCC Mercator Research Institute on Global Commons and Climate Change

Technische Universität Berlin

Contact

Email: nachtigall@tu-berlin.de

Website: floriannachtigall.github.io

*Appendix*

# Study area

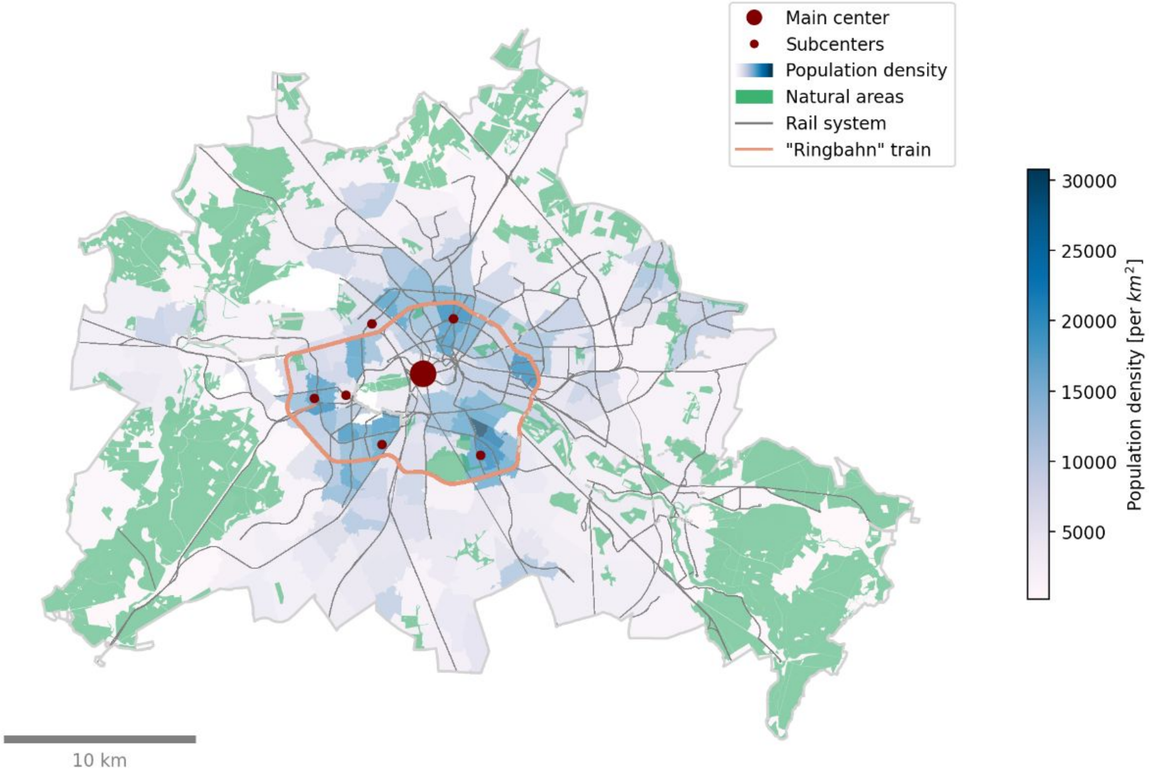



**Figure 4:** ***Built environment of Berlin, Germany.*** *The center and subcenters, based on points of interest density, are indicated as dark red circles. Population density of neighborhoods is color coded in blue. Natural areas according to Berlin land use data [20] are marked in green. The public 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.*

*Appendix*

# Travel behavior

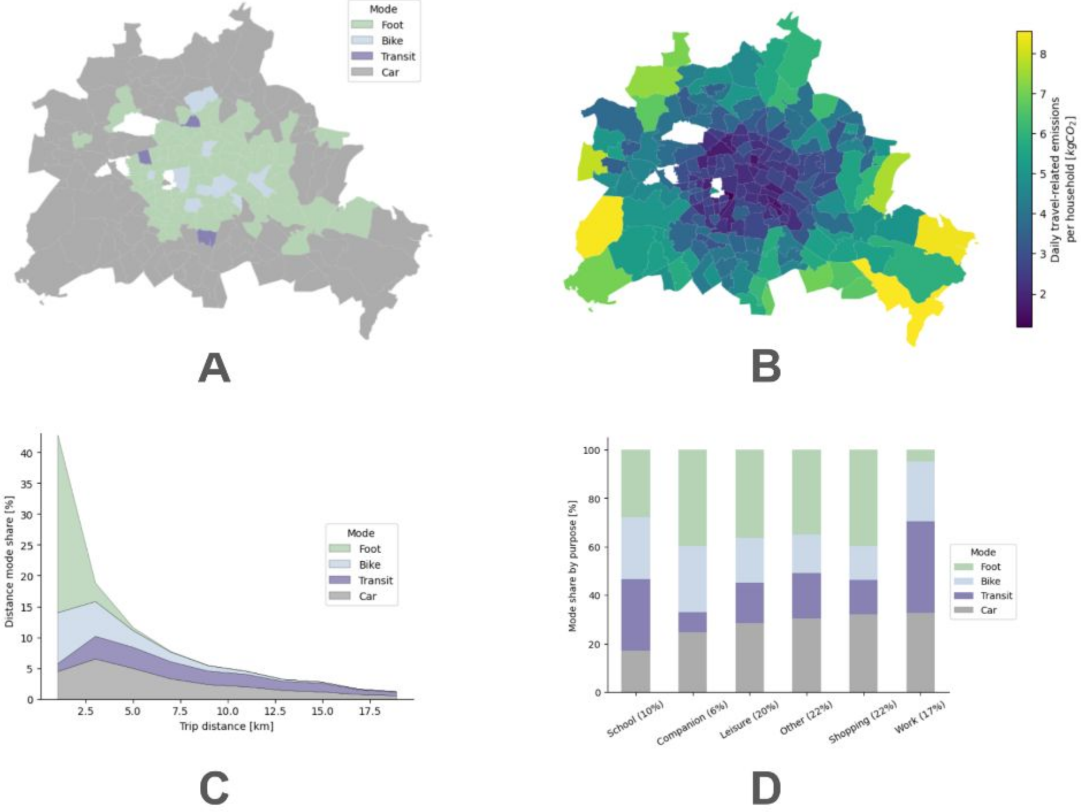



**Figure 3:** ***Overview of trip mode, purpose, distance, and related $CO_2$ emissions.*** *(A) Predominant mode of transport for each neighborhood based on trip counts. (B) Average travel-related $CO_2$ emissions per household for each neighborhood. (C) Trip distance specific mode share. (D) Trip purpose specific mode share, ordered by increasing car share.*

*Methods*

# Feature engineering: Built environment & travel preferences

| 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 mix-use areas |
| Design | Car-friendliness index | Provision of expressway kilometers per capita |
| | Walkability index | Intersection density in the built-up area |
| Distance to transit | Transit accessibility index | Gravity model-based index describing the average spatio-temporal transit accessibility of a neighborhood |

**Table 2:** ***Built environment characteristics.*** *Overview of all built environment characteristics considered in*

| Category | Variable name | Description |
|---|---|---|
| Socio-demographics | Income | Average household income |
| | Household size | Average number of persons living in a household |
| | Age | Average age of adult (>18 years) residents |
| | Higher education | 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) |
| | Political preferences | Electoral share of the Green party in constituencies intersecting the neighborhood in the last regional elections |

**Table 1:** ***Travel preferences.*** *Overview of all socio-demographic traits and proxies for travel-related attitudes*

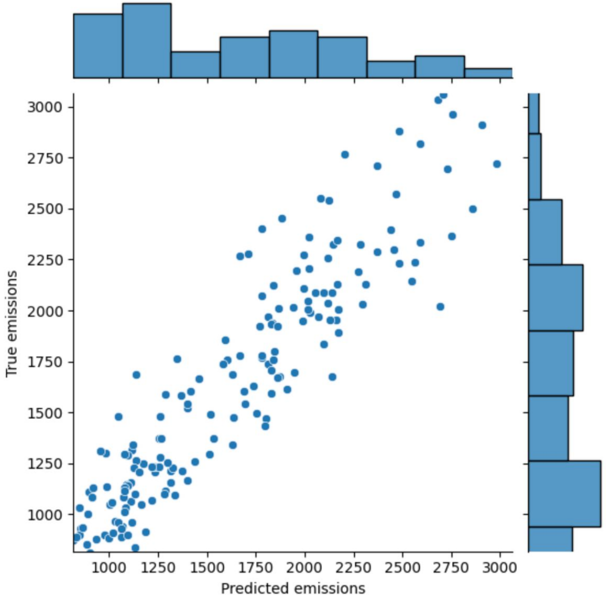
# Emission factors to convert travel distance to $CO_2$ emissions

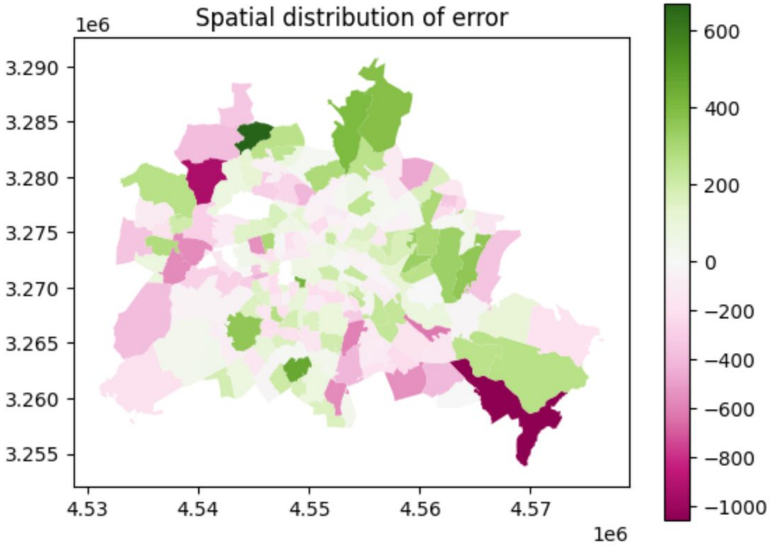
| Mode | Emissions [g $CO_2$/pkm] |
|---|---|
| Car (ICE) | 162 |
| Moped (ICE) | 70 |
| Transit | 65 |
| Bike | 20 |
| Foot | 0 |

**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 $CO_2$ 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.*

# Explanatory power of covariates with XGBoost regressor

- 5-fold random cross-validation with 1000 tree estimators, a tree depth of 6, and a learning rate of 0.01
- Coefficient of determination, $R^2$, between 0.8 and 0.85 depending built environment characterization and inclusion of transport means ownership attributes



*Appendix*

# Causal inference: Double machine learning

**GOAL**

- Estimate causal effects from observational data

**WHY**

- Randomized control trial (RCT) is not suitable
- Confounding effects (treatment assignment is not randomized, leading to biased estimates)
- High-dimensional covariates (functional form unknown or non-parametric)
- Multiple, continuous treatment dimensions

**HOW**

- Stage 1: Debiasing / estimation of nuisance parameters
  - Predicting the outcome from the controls -> *outcome residuals*
  - Predicting the treatment from the controls -> *treatment residuals*
- Stage 2: Estimation of heterogeneous treatment effect
  - Predicting the *outcome residuals* from the *treatment residuals* and controls

**MODEL**

- XGBoost for nuisance models, CausalForest for final model (EconML CausalForestDML implementation)