Journal of Transport Geography 100 (2022) 103307

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Ride-pooling demand prediction: A spatiotemporal assessment in Germany

Felix Zwick[a,b,*], Kay W. Axhausen[a]

[a] IVT, ETH Zurich, CH-8093 Zurich, Switzerland
[b] MOIA GmbH, Hamburg, Germany

ARTICLE INFO

Keywords:
Ride-sharing
On-demand mobility
Spatial regression
Open data
New mobility
Demand prediction
Ride-splitting

ABSTRACT

Ride-pooling has attracted considerable attention from both academia and practitioners in recent years, promising to reduce traffic volumes and its negative impacts in urban areas. Simulation studies have shown that large-scale ride-pooling has the potential to increase vehicle utilization, thereby reducing vehicle kilometers traveled (VKT) and required fleet sizes compared to single-passenger mobility options. However, in the real world, large-scale ride-pooling services are rare and not yet widely implemented, in part due to high operating costs that are expected to decrease substantially with the advent of automated vehicles.

Two large-scale ride-pooling fleets are operated by MOIA in Hamburg and Hanover, Germany, and serve as testbeds for future pooled services. For this study, we analyze pre-pandemic demand data from both services from 2019 and 2020 and perform spatial and random forest regressions to understand the (spatial) characteristics of ride-pooling trip origins in both cities. We then examine how well findings from one study area (Hamburg in our case) can be generalized and transferred to other cities (Hanover in our case) to enable spatial predictions beyond areas with an existing service.

The regression results are similar in both cities and show the strongest impact on ridepooling demand by the variables capturing the density of workplaces, gastronomy and culture. Given the regression results for Hamburg, we predict trip origins for Hanover and observe that demand is overestimated by all applied models. The spatial lag of X (SLX) model showed the most promising results with an overall overestimation of trip origins below 20%.

1. Introduction

Ride-pooling systems have gained a lot of attention in academia and practice in recent years and promise flexible and convenient mobility that is more efficient than private car traffic or ride-hailing services. Multiple passengers with a similar route are transported with only one vehicle and thus VKT, vehicle fleets and negative externalities like congestion or noise can be reduced (Martinez and Viegas, 2017; Alonso-Mora et al., 2017; Greenblatt and Saxena, 2015; Zwick et al., 2021b). Customers face slight detours, wait times and potential walks to the virtual or physical stops of the system. A similar form of mobility already has existed for decades in an unregulated, privately organized analog form in emerging economies, e.g. the minibus (or jitney) system in South Africa (Neumann et al., 2015) or colectivos in Chile (Zwick, 2017), filling a gap in public transportation. Routes and schedules are semi-flexible in these cases and not transparent to passengers, leading to inefficiencies and a lack of information for travelers.

In comparison, emerging digitized on-demand mobility providers, Transportation Network Companies (TNCs), offer fully flexible services without pre-defined routes or schedules and are currently dominated by ride-hailing services like Uber (2021), Lyft (2021) or DiDi (2021) that mainly operate in North and South America and Asia. Many simulation studies showed that the advent of ride-hailing platforms lead to an increase in vehicle kilometers traveled (VKT) in urban areas due to deadheading, induced trips and cannibalization of more sustainable modes (Tirachini and Gomez-Lobo, 2020; Henao and Marshall, 2019; Pernestål and Kristoffersson, 2019; Jing et al., 2020).

In addition to their ride-hailing business, the named TNCs also offer

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; GHG, greenhouse gas; GWR, geographically weighted regression; GWPR, geographically weighted poisson regression; MAE, mean absolute error; MTS, main train station; OSM, OpenStreetMap; POI, point of interest; RENB, random-effects negative binomial; RF, random forest; RMSE, root mean square error; SEM, spatial error model; SLX, spatial lag of X; SD, standard deviation; SDEM, spatial Durbin error model; TNC, transportation network companies; VKT, vehicle kilometers traveled.

* Corresponding author at: IVT, ETH Zurich, CH-8093 Zurich, Switzerland; MOIA GmbH, Hamburg, Germany.
E-mail address: felix.zwick@ivt.baug.ethz.ch (F. Zwick).

https://doi.org/10.1016/j.jtrangeo.2022.103307
Received 3 October 2021; Received in revised form 17 February 2022; Accepted 18 February 2022
Available online 3 March 2022

pooled services (i.e. ride-pooling). However, the share of pooled rides was found to be below 30% (Li et al., 2019; Chen et al., 2021; Soria et al., 2020; Young et al., 2020; Tirachini, 2020; Brown, 2019; Wang and Noland, 2021b). During the pandemic, the option to book pooled rides was even deactivated by Uber and Lyft to avoid virus spreading (Kucharski et al., 2021) and has been reactivated by Lyft only in July 2021 (Bellon, 2021).

Due to the potential negative impact on traffic and stricter legislation to protect traditional cab services, ride-hailing services are subject to strict regulations and are less common in Europe and especially in Germany (Cetin and Deakin, 2019; Ennen and Heilker, 2020). Recently, German legislation has been amended to allow on-demand services that exclusively offer pooled trips (Ritzer-Angerer, 2021). MOIA is an operator of exclusive ride-pooling services and has been operating its service in Hamburg and Hanover since 2019 and 2017, respectively, under an experimental clause.

In this study, we use pre-pandemic MOIA trips and estimate (spatial) regression models to understand the spatial characteristics of ride-pooling trip origins in the two cities. We investigate, how well the findings can be transferred from one city to the other. Given the model parameters of Hamburg, we predict potential ride-pooling trip origins in Hanover and compare the prediction to real-world observations.

In this way, we contribute to the existing literature with a detailed empirical investigation of revealed data from a European large-scale ride-pooling service that exclusively offers pooled trips. The demand prediction allows insights into how well the findings can be transferred and what limitations exist.

1.1. Ride-pooling services and recent research findings

Ride-pooling services have emerged rapidly in recent years. Foljanty (2020) mapped on-demand ride-pooling services worldwide and found more than 300 services, exclusively offering ride-pooling. TNCs like Uber or Lyft are not included as their main business is ride-hailing. Although the list of services is likely incomplete and quickly evolving, it provides an informative overview of the current market penetration of ride-pooling. On average, the services operate with 20 vehicles, and large-scale services with more than 100 vehicles are scarce.

Simulation studies, however, showed that the pooling potential unfolds only with large-scale services with a high demand and vehicle density (Engelhardt et al., 2019; Kaddoura and Schlenther, 2021; Zwick et al., 2021c). Engelhardt et al. (2019) state that in Munich, Germany, an adoption of at least 5% of private vehicle trips and 1000 vehicles are required to allow a reduction of VKT. Zwick et al. (2021c) found a logarithmic relationship between trip density and service efficiency and observed that average vehicle occupancy is more than doubled when comparing a large-scale and a small-scale service. The importance of a large share of pooled trips is also shown in a study in Chengdu, China, by Li et al. (2021) where the potential to reduce greenhouse gas (GHG) emissions through pooling is estimated to be roughly 30%.

Other simulation studies focused only on large-scale ride-pooling systems and reported many beneficial impacts on the transport system (Zwick et al., 2021b; Martinez and Viegas, 2017; Engelhardt et al., 2019; Alonso-Mora et al., 2017). The main benefits over conventional private car traffic are reduced vehicle fleets and VKT. Martinez and Viegas (2017) found a VKT reduction of 25% and a vehicle replacement rate of 1:20 when all private car, bus and taxi trips in Lisbon are replaced with ride-pooling. Zwick et al. (2021b) simulated four scenarios and found that one ride-pooling vehicle can transport between 63 and 153 passengers per day (dependent on penetration rate and service design), which is substantially higher than with private vehicles and would allow substantial reduction of vehicle fleets. VKT are reduced by up to 54% if accompanying measures are taken to ban conventional car traffic.

The necessity of accompanying policies when autonomous vehicles and on-demand service are introduced is also pointed out by Pernestål and Kristoffersson (2019) and Naumov et al. (2020). They argue that such policies may promote higher pooling rates and also avoid a cannibalization of public transportation. Their findings suggest that ride-pooling is, despite lower fares than for ride-hailing, only under certain circumstances attractive enough to pull travelers from private cars and hailing services. Ruch et al. (2020) simulated an urban and a rural pooled service and found efficiency gains being relatively small, so that it is questionable if they compensate the loss of privacy and time for customers. However, only 250 vehicles were employed in the largest considered service and reduced congestion for other transport modes were not considered. Ke et al. (2020) modelled the time losses for ridesourcing customers and car drivers in a theoretical scenario and found win-win situations for both car drivers and ridesourcing customers when passenger demand is high enough (at least 14,000 trips/h) and pooling enabled.

There has been extensive research on the users and travel behavior with regard to on-demand mobility services. Tirachini (2020) provides a detailed review of conducted surveys and of research into trip and GPS data, mainly from ride-hailing services. We concentrate on research on pooled on-demand services here.

Chen et al. (2017) estimate a machine learning model to find the main factors for the choice of ride-pooling over ride-hailing in Hangzhou, China. The most important factors are travel time, surge pricing rate (for ride-hailing), trip fares and trip distance. Morales Sarriera et al. (2017) and de Souza Silva et al. (2018) conducted surveys on the use of ride-hailing and ride-pooling in Brazil and the US, respectively. They found that, apart from the fare, the (perceived) safety and information about future passengers (in terms of race and gender) in the vehicle is a critical factor to choose a pooled ride. A high impact of the fare on the ride-pooling choice is also found by Wang and Noland (2021b). A survey analysis in the Greater Boston region by Gehrke et al. (2021) showed that respondents with a lower income are more likely to opt for a pooled trip. The same accounts for people younger than 25 years and people identifying themselves as Black or African American. Spurlock et al. (2019) confirm that low- and middle-income cohorts are more likely to use ride-pooling over ride-hailing and argue that the pooled service facilitates the access to the benefits of emerging mobility services. All the reported findings on choices of ride-pooling over ride-hailing show a clear trend that income and fares are the most decisive factors for or against choosing a pooled ride.

Respondents of the survey of Kostorz et al. (2021) in Hamburg, Germany, had no option to choose ride-hailing since it is barely an alternative in Germany due to previously mentioned legal restrictions. The study revealed that, in contrast to the previously discussed studies, ride-pooling users have a higher income than non-users. 60% of the respondents reported that their last trip purpose was *leisure*, which is in line with results from another survey in Germany by Knie and Ruhrort (2020).

1.2. Studies on spatial demand characteristics of on-demand mobility

Multiple studies analyzed the spatiotemporal characteristics of on-demand mobility trips. In the following an overview is given, starting with ride-hailing services, continuing with ride-pooling services that are offered as an alternative to ride-hailing and concluding with services exclusively offering ride-pooling.

Yu and Peng (2019) employed a spatial error model (SEM) and a geographically weighted poisson regression (GWPR) to analyze the spatial variation of ride-sourcing/ride-hailing demand and its relationship to the built environment and socioeconomic factors in Austin, Texas. They found a general positive impact on ride-hailing demand with increasing population, road and sidewalk density, land-use diversity and transit accessibility, whereas for areas with a higher median age of the population and a better balance between population and employment they found a lower demand. Furthermore, they found that the income impact varies on weekdays and weekends. Ghaffar et al. (2020) also found a higher demand in densely populated areas using a random-effects negative binomial (RENB) regression approach to analyze determinants of ridesourcing usage in Chicago. Additionally, a higher demand was found in census tracts with higher incomes, higher employment density and higher land-use diversity among others. A cross-city analysis is provided by Sabouri et al. (2020) who had access to Uber trip data from 24 US regions. They used a hierarchical linear model to explore ride-hailing demand characteristics and found a positive impact of population, employment, activity density, land use entropy and transit stops on ride-pooling demand. Wang and Noland (2021a) present similar findings applying an ordinary least squares (OLS) model and a geographically weighted regression (GWR) model on demand data of the TNC Didi in Chengdu, China. According to their study population and local road density, housing prices and proximity to subways have a general positive impact on the number of trips.

Summing up, the studies discussed indicate that spatial characteristics of ride-hailing demand are positively influenced by population and activity density, higher incomes and land-use diversity.

The following studies analyzed on-demand services that offer both ride-hailing and ride-pooling and identified spatial determinants of ride-pooling use.

Li et al. (2019) used demand data of Didi in Chengdu, China, and compared the spatiotemporal patterns of pooled and unpooled trips. Pooled trips occur at later times of the day than unpooled trips, which indicates a lower attractiveness for commuting trips. Furthermore, pooled trips more often start in the city center, while unpooled trips more often start outside the city center.

The share of pooled trip has also been analyzed by Hou et al. (2020) for data from the TNCs Uber, Lyft and Via in Chicago. According to their study the most important predictors of the (un-)willingness to pool are identified to be the income level and airport trips, which both have a negative effect on the share of pooled trips. Brown (2019), Malik et al. (2021) and Dean and Kockelman (2021) revealed for Los Angeles, California and Chicago, respectively, that the pooled hailing option is more likely to be selected by younger individuals and individuals with low incomes. This confirms the findings from several surveys. Soria and Stathopoulos (2021) confirmed the positive relation between socio-economic disadvantages and ride-pooling usage and additionally found a higher rail transit access to be associated with higher ride-hailing and ride-pooling usage.

Brown (2019) and Gehrke et al. (2021) also found that in areas of high population density in Los Angeles and Boston, respectively, there is a greater likelihood that a pooled service option will be chosen.

Most of the existing studies consider ride-pooling services that are offered additionally to the hailing option since exclusive ride-pooling services are scarce. However, few studies exist for the ride-pooling landscape in Germany:

Aberle (2020) analyzed the service area characteristics of four on-demand mobility providers in Hamburg. He found that three out of four service areas mainly cover the central areas with higher incomes. However, the author had no demand data available and only considered the supply side.

Knie and Ruhrort (2020) plotted trip origins and destinations of the ride-pooling service CleverShuttle (CleverShuttle, 2021) in Berlin, Dresden, Munich and Leipzig and found a concentration of trips around the city centers. No further spatial investigations were made.

Another spatial regression study on the MOIA service in Hamburg considers three time periods before and during the COVID-19 pandemic (Zwick et al., 2021a). This study shows that ride-pooling demand is significantly higher in areas with a high number of inhabitants, workplaces, rail stops and gastronomic facilities, in areas with inhabitants having a high social status or welfare, and at the airport. The impact is similar for trip origins and destinations. The overall picture has been shown to remain similar during the pandemic, but the influence of gastronomy and the airport has decreased.

1.3. Study objectives and contributions

In this study, we use trip data of the MOIA services in Hamburg and Hanover from May 2019 to February 2020 for multiple regression models to understand the relationship between spatial data and demand. Based on previous investigations that are presented and discussed in Section 2.1, we decided to use *trip origins* as the demand variable. We estimate a general OLS model, a spatial lag of X (SLX) model (Halleck Vega and Elhorst, 2015) and a random forest (RF) regression model (Breiman, 2001), which gives us a comprehensive understanding of the demand patterns in both cities.

We then compare the demand patterns and use the Hamburg data to predict ride-pooling demand in Hanover based on the spatial structure. The prediction accuracy is quantified for each model and it is discussed which model is best suited for the task at hand.

A similar methodology was applied in previous studies (Guidon et al., 2020; Dean and Kockelman, 2021), although SLX and RF models are usually less common.

Our study may support transport planners, operators and policy makers identify the main drivers of ride-pooling demand and estimate the potential spatial distribution of demand in new cities, only based on spatial data that is typically available for any city in developed countries. This enables informed planning of a new ride-pooling service, e.g., with respect to the shape of the service area, the number of vehicles and the number of drivers needed, the distribution of depots, or the setting of the vehicle repositioning algorithm.

The contributions to the literature are as follows:

- The investigation adds to the growing body of empirical evidence on ride-pooling usage and its spatiotemporal characteristics.
- We provide a comprehensive analysis of the largest European contiguous ride-pooling service in two cities and identify the most relevant impact factors of its demand.
- Three regression models are applied and evaluated with respect to their model quality and predictive accuracy.
- The transferability and prediction accuracy of the service's demand to new cities is evaluated and quantified.

2. Dependent and independent input variables

We use input data from multiple sources. The demand data were obtained from the ride-pooling operator MOIA. The spatial data were provided by the cities of Hamburg and Hanover, and additionally public OpenStreetMap (OSM) (OpenStreetMap Contributors, 2019) data were used.

2.1. Ride-pooling demand data

Two of the largest comprehensive ride-pooling systems in Europe are operated by the company MOIA in Hamburg and Hanover. MOIA was founded in 2016 and established the first ride-pooling services in Hamburg and Hanover in 2019 and 2017, respectively. In Hamburg, MOIA operates with up to 500 vehicles, which is Europe's largest ride-pooling fleet in a single city. The services are a testbed for future digital mobility services, which are expected to grow in the future.

There are multiple simulation studies on the implications of MOIA on Hamburg's transport system (Zwick and Axhausen, 2020a, 2020b; Hörl and Zwick, 2021; Kagerbauer et al., 2021; Kuehnel et al., 2021; Wilkes et al., 2021). Zwick et al. (2020), Kagerbauer et al. (2021) and Kostorz et al. (2021) analyzed users, usage patterns and adapted mobility behavior due to the newly available mode. We use MOIA trip data of Hamburg and Hanover for the time between May 2019 and February 2020, meaning that demand was not yet influenced by the COVID-19 pandemic. A more recent analysis by Zwick et al. (2021a) showed that the spatial distribution of trips only changes gradually with the pandemic and the overall picture remains similar. However, the impact of COVID-19 on the demand needs to be further explored in future.

Besides trips, one could also use MOIA *requests* as a demand indicator for ride-pooling. We also considered requests for this study and found a very strong correlation with rides (correlation coefficients in Hamburg/ Hanover: 0.98/0.99). Therefore, the model results are not expected to differ much when the demand variable is changed. We preferred trips to requests for these reasons:

- It is unclear whether the intention of a request was to ride with MOIA or just to check out a possible alternative. Especially in the initial phase of MOIA, many requests were observed without a ride intention.
- Requests are often submitted multiple times, especially when rejected by the system. Therefore, a single trip can result in many submitted requests. The resubmissions could be identified and purged, but this would require a manual and to some degree arbitrary purge process.
- Trips are actually revealed preferences and are unlikely to be misinterpreted.

We observed 1,228,530 trips in Hamburg and 330,634 trips in Hanover during the time period. One trip can include up to six passengers. The temporal distribution of the trips is shown in Fig. 1. Note that relative shares are shown and the absolute numbers are higher for the service in Hamburg.

The share of all trips in each considered month (a) shows an increase of trips over time. Since the Hamburg service only started its operation in April 2019, the numbers reflect the ramp-up phase of demand and

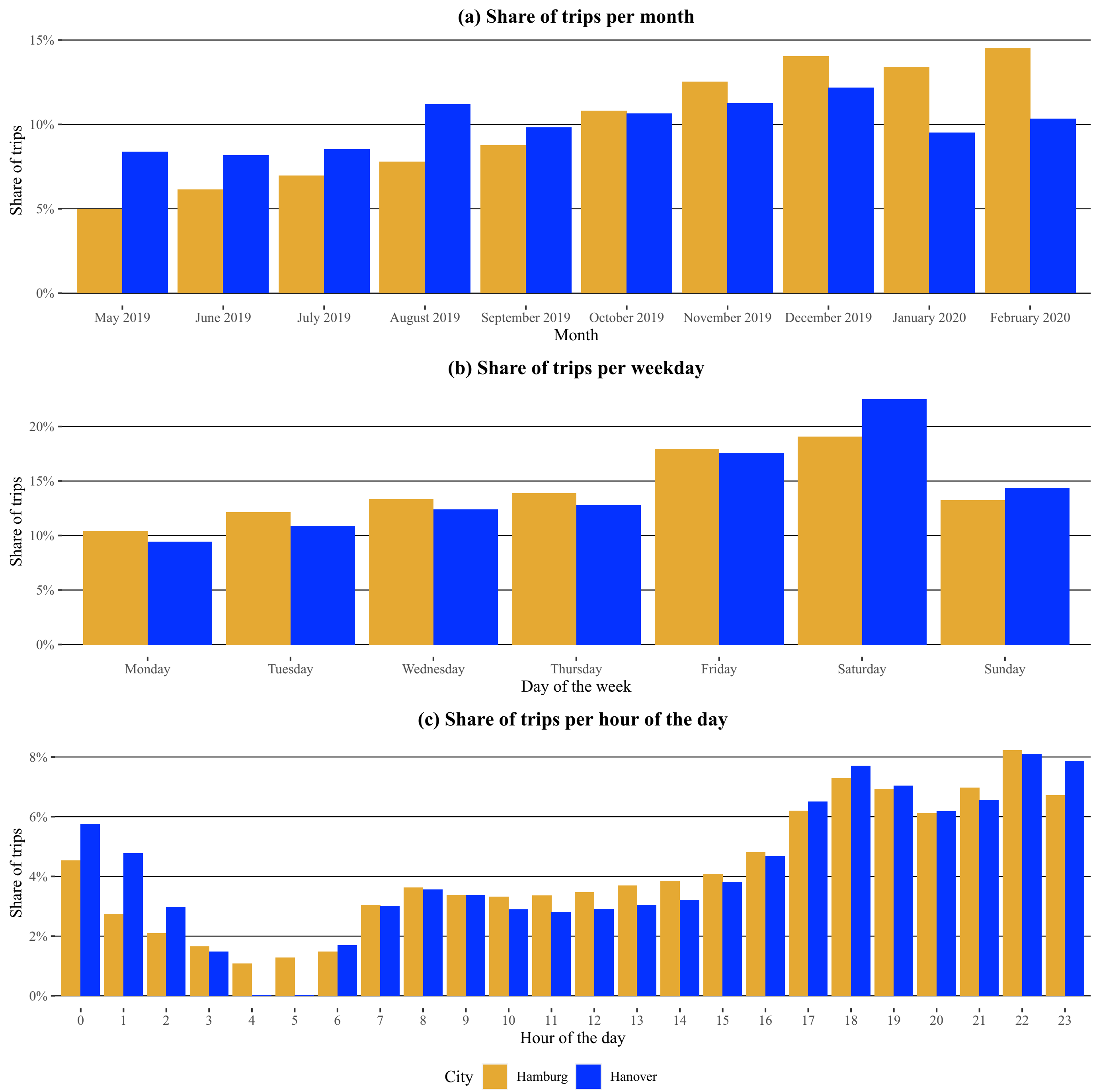

Fig. 1. Temporal MOIA demand distribution in Hamburg (yellow) and Hanover (blue).

supply. While the Hamburg service started its operation with only 100 vehicles, there were more than 300 vehicles operating by the beginning of 2020.

The share per weekday (b) shows that demand is lowest on Mondays and increases slightly in the course of the week, peaking on Fridays and Saturdays. The Saturday peak is higher in Hanover, but the overall distribution in both cities is otherwise very similar.

The trip distribution over the course of a day (c) also shows similar patterns for both cities. Due to limited service times at night in Hanover, there was no demand between 4 am and 6 am. There are two demand peaks, one in the early evening between 6 pm and 8 pm and one in the late evening between 10 pm and midnight.

Fig. 2 shows the spatial distribution of ride-pooling trip origins of MOIA in Hamburg and Hanover for the period from May 2019 to February 2020. They are mapped to 650 statistical zones in Hamburg and 389 statistical zones in Hanover and serve as the dependent variables in our investigation. The Hamburg service area covers the northern part of the city and has an overall size of 195 km^2 whereas the Hanover service area covers the entire city area of 204 km^2.

The spatial distribution of origins and destinations show similar patterns (Zwick et al., 2021a). For simplicity, we only focus on trip origins here. Trips that were requested from outside the service areas and included a major walking stage to the service area were not further considered, which reduces the number of considered trips to 1,106,338 in Hamburg and to 301,399 in Hanover.

Demand in Hamburg is concentrated in the city center in the south of the service area, where the main train station is also located. Another demand cluster is the airport in the north of the city. Areas with very low demand are mainly located in parks or green areas.

In Hanover the overall demand level is lower than in Hamburg. It is also concentrated in the city center and decreases toward the edges of the service area. Low-demand areas are located primarily in the northern and eastern parts of the city, which are also less populated. The Hanover airport is outside the service area, as MOIA is only allowed to operate within the city limits.

2.2. Spatial data

The spatial data were obtained from multiple sources. The cities of Hamburg (2018) and Hanover (2020) kindly provided the statistical zones and up-to-date population, employment and car ownership data. Additionally, we extracted OSM data from November 2019 to identify points of interest (POIs).

The following data were used, sorted by source:

Hamburg (2018) and Hanover (2020):

- Number of inhabitants.
- Number of workplaces (Hanover data from 2019).

OpenStreetMap Contributors (2019):

- Gastronomy: Amenities of the category *sustenance*.
- Culture: Amenities of the category *entertainment, arts & culture*.
- Shops: Shops of the category *general store, department store, mall*.
- Rail stops: Number of railway stops, weighted with the number of platforms.

It should be noted that specifically the OSM data are subject to statistical noise and do not have a uniform impact. For instance, a large restaurant is more likely to attract ride-pooling trips than a small restaurant. Another example is the rail stop network in both cities: Whereas in Hamburg the subway and suburban trains always run separately from road traffic, in Hanover rail-bound public transport runs partly parallel to road traffic and is more widespread.

We additionally introduced a centrality variable *distance to main train station* (distance MTS) by calculating the distance of each cell's centroid to the respective city's main train station to reflect the distance to the city center. This variable may capture the impact of trip-generating activities and accessibilities associated with city-specific land-use patterns, which are omitted by the other independent variables (Guidon et al., 2020). Since we aim to understand the explanatory power of this variable, we estimate each model with and without the centrality variable.

In Hamburg, the airport is also part of the service area. However, the airport's attraction is largely caused by a single factor, the airport itself, which is why we excluded the airport area from the model estimation. We additionally investigated the impact of hospitals, clothing shops, average age and car ownership. Since they did not substantially improve the models and they mostly showed no significant impact on demand, we decided to exclude these items from the final models.

While the data provided by the cities were already mapped to the statistical zones provided, the OSM data were reviewed, duplicates were removed, and the data were then matched to the statistical zones.

Table 1 provides a summary of the variable distribution in the statistical zones including the minimum, the first quartile, median, mean, standard deviation (SD), third quartile and maximum.

3. Methodology

We used multiple regression models to understand the relationship between trip origins and the presented spatial variables. First, we estimated a global OLS model to get an overall impression of each variable's impact. Given the OLS model and the spatial structure, defined by a neighborhood matrix, we determined the most suitable spatial regression model through multiple statistical tests. Finally, we also performed a random forest regression to account for non-linear relationships.

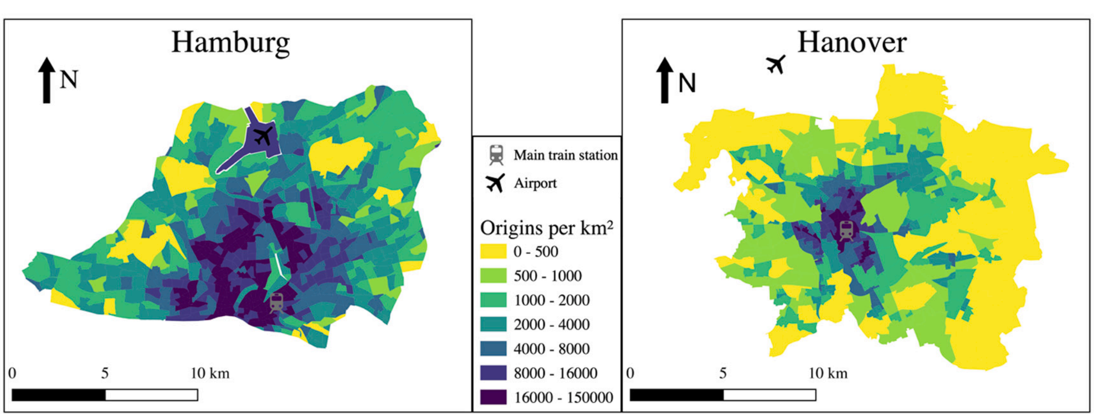

Fig. 2. Spatial MOIA demand distribution in Hamburg (left) and Hanover (right) for the period from May 2019 to February 2020. Hanover: basic data © Landeshauptstadt Hannover, Geoinformation.

Table 1
Overview of cell-based demand and spatial data distributions.

		Min.	1st quart.	Median	Mean	SD	3rd quart.	Max.
Trip origins [× 1000]	Hamburg	0.0	0.5	1.1	1.8	2.6	2.3	45.7
	Hanover	0.0	0.3	0.5	0.8	1.1	0.9	14
Population size [× 1000]	Hamburg	0	0.8	1.9	1.7	1.1	2.5	4.8
	Hanover	0.9	1.2	1.4	1.4	0.2	1.6	2.6
Jobs [× 1000]	Hamburg	0	0.2	0.5	1.5	2.5	1.7	18
	Hanover	0	0.1	0.3	0.9	2.4	0.8	32.4
Shops POIs	Hamburg	0	0	1	1.7	2.5	2	17
	Hanover	0	0	1	1.8	3.0	2	38
Gastronomy POIs	Hamburg	0	1	3	6.7	11.2	7	84
	Hanover	0	1	2	4.4	8.3	5	111
Culture POIs	Hamburg	0	0	0	0.8	2.3	1	25
	Hanover	0	0	0	0.8	2.0	1	19
Rail stops	Hamburg	0	0	0	0.4	1.2	0	18
	Hanover	0	0	0	1.0	1.9	1	19
Dist. to main station [km]	Hamburg	0	3.5	5.5	5.8	3.0	7.9	13.4
	Hanover	0.3	2.3	4.2	4.2	2.1	5.7	9.7

All models were estimated for Hamburg and Hanover. We then used the estimated Hamburg models and the spatial data from Hanover to predict the demand for ride-pooling in Hanover and compared the predicted trip origins with the observed trip origins.

3.1. Linear and spatial regression

The OLS model is a common tool to estimate the relationship between one or multiple independent explanatory variables and a dependent variable (*ride-pooling trip origins* in our case). The model is defined by

$$y = X\beta + \varepsilon, \tag{1}$$

where y denotes the dependent variable, X denotes the independent variable(s), β denotes an estimated coefficient and ε denotes an error term.

As the OLS model does not take into account any spatial autocorrelation within the data, we additionally applied a spatial regression model, for which we defined rook contiguity-based weights that describe all neighboring zones of each zone. In Hamburg, we removed connections between zones across the lake *Alstersee*.

With the OLS model and the neighboring matrix, we used Moran's I test (Moran, 1950) and Lagrange-Multiplier tests (Anselin, 1988a) to detect spatial autocorrelation. In our case the Moran's I test indicates autocorrelation for all models (p-value <0.01), which is a strong indicator that a spatial regression model explains the data better than the OLS model.

To identify the most suitable spatial regression model, we performed Lagrange-Multiplier tests. The tests indicate a spatial lag and a spatial error dependence for all models. Thus, we initially considered two main classes of spatial regression models to identify the most suitable model: models with a lag on Y (the dependent variable), *global spatial models*, and spatial error models. We assume that the dependent variable does not cause a spatial lag because ride-pooling demand is not affected by ride-pooling demand in neighboring zones, but rather by the independent variables in the neighboring zones. Therefore, we discarded the models with a lag on Y. We additionally considered models with a lag on X (the independent variables), which in the literature are typically referred to as *local spatial models*.

We followed the approach of Anselin (1988b) and LeSage and Pace (2009) and estimated a spatial error model (SEM), a spatial lag of X model (SLX), and a spatial Durbin error model (SDEM) that combines a spatial lag *and* a spatial error in one model. The model fit for Hamburg and Hanover is best with the SDEM. However, the prediction turned out to be worse with the models that contain a spatial error, which indicates that the spatial error cannot be transferred well across cities. Thus, we decided to apply the SLX model, which is defined by

$$y = X\beta + WX\theta + \varepsilon, \tag{2}$$

where W denotes the spatial weight matrix and θ denotes the coefficient for local spatial spillovers.

3.2. Random forest

The random forest regression is a machine learning approach for classification and regression tasks (Breiman, 2001). Multiple decision trees are built up on the training dataset with the bootstrap aggregating (Bagging) method and used as new learning sets (Breiman, 1996). The results are averaged to ensure a stable result and to minimize noise and the risk of overfitting the model.

The method can be used to assess the relative importance of each variable by evaluating how much the trees including the variable reduce impurity across all trees. The variable importance is an essential measure as it hints to variables that can be excluded and are less important for prediction.

Additionally, partial dependencies between independent variables and the dependent variable can be plotted. The plots show marginal effects that give a valuable overview of the data pattern and distribution.

4. Results

In the following we present the OLS, the SLX and the RF regression model results. For each model, we estimate two model versions, one without the centrality variable *distance to main train station* and one with. Thus, we overall estimate six models for each city. The Hamburg parameters are then used for the demand prediction for Hanover to evaluate how well it matches reality. The most suitable prediction model is identified and further explored.

4.1. (Spatial) regression models

The OLS and SLX regression results for Hamburg are shown in Table 2. For each model, a direct effect of the independent variable on ride-pooling trips is given. For the SLX model, an additional indirect spillover effect is given, the spatial lag X. This parameter shows the impact on demand in neighboring zones of a zone when the independent variable in a zone is increased. The two coefficients of determination Multiple R^2 and Adjusted R^2 as well as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are given.

All models have similar coefficients of determination (R^2) varying from 0.67 to 0.72. For the OLS models, all selected variables have a significant impact. All effects are positive except the number of shops. The population size seems to be more important for a high ride-pooling demand than the number of *jobs* in a cell. The number of *rail stops* and

Table 2
Linear and spatial regression results for Hamburg.

	Hamburg			
	OLS	SLX	OLS	SLX
			Incl. centrality variable	
Intercept	93.90	−184.31	738.31***	697.54***
Population	376.56***	346.13***	388.25***	344.32***
Jobs	96.90***	90.42***	73.85***	82.51***
Rail stops	171.19***	131.58***	153.60***	125.90***
Dist. MTS			−98.75***	−289.59
Shops	−72.22**	−63.88**	−67.30**	−63.85**
Culture	208.01***	201.24***	195.86***	199.93***
Gastronomy	104.50***	104.14***	97.15***	102.47***
Lag pop.		260.37**		256.08**
Lag jobs		91.64*		32.99
Lag rail St.		−361.28***		−390.97***
Lag dist. MTS				179.72
Lag shops		−151.22**		−159.36***
Lag culture		−66.55		−77.45*
Lag gastron.		39.61***		28.72**
Multiple R^2	0.67	0.70	0.69	0.72
Adjusted R^2	0.67	0.69	0.68	0.71
AIC	10,996	10,943	10,962	10,908
BIC	11,031	11,006	11,003	10,980

Note: ***p-value < 0.001; **p-value < 0.01; * p-value < 0.05;. p-value < 0.1.

Table 3
Linear and spatial regression results for Hanover.

	Hanover			
	OLS	SLX	OLS	SLX
			Incl. centrality variable	
Intercept	81.57	255.04	437.99**	474.38
Population	149.33	118.92	157.83	122.97
Jobs	104.12***	104.00***	111.47***	107.28***
Rail stops	20.10	17.98	35.15*	25.82
Dist. MTS			−83.29***	−163.47
Shops	−42.98***	−39.34**	−43.39***	−38.18**
Culture	130.59***	113.66***	106.58***	105.67***
Gastronomy	78.85***	70.55***	73.46***	69.91***
Lag pop.		−68.71		−9.92
Lag jobs		65.31**		70.40**
Lag rail st.		−57.15		−18.44
Lag dist. MTS				93.50
Lag shops		−98.24***		−93.13***
Lag culture		134.54***		84.59*
Lag gastron.		14.55		5.83
Multiple R^2	0.78	0.80	0.80	0.81
Adjusted R^2	0.77	0.79	0.80	0.81
AIC	5980	5952	5942	5929
BIC	6011	6007	5978	5993

Note: ***p-value < 0.001; ***p*-value < 0.01; *p-value < 0.05;. p-value < 0.1.

POIs of the category *culture* have a higher impact than POIs of the category *gastronomy*, which is not surprising given that the overall quantity of *gastronomy* facilities is much higher than *rail stops* or *cultural* facilities.

Including the centrality variable *Distance MTS* leads to a slightly better model fit. The variable is significant and has, as expected, a negative sign in the OLS model.

The SLX model shows similar direct impacts. The indirect (spillover) impacts for *population* and *jobs* are positive and significant. The indirect effect of *rail stops* is negative, which indicates that although *rail stops* have a positive direct impact on demand, the areas around train stations generate fewer trips. *Shops* and *cultural* facilities also have a negative indirect impact. The indirect impact of *gastronomy* is positive and indicates a general attractiveness of areas with a proximity to bars and restaurants.

Including the centrality variable again leads to a slightly better coefficient of determination and the overall pattern remains. The centrality variable however does not have a significant direct or indirect impact on demand in the SLX version.

The results of the same models applied to Hanover are shown in Table 3.

The R^2-values are even higher than for the Hamburg models and vary from 0.77 to 0.81. The signs of all variables in the OLS model are the same as in the Hamburg model. However, surprisingly the positive impact of *population* and *rail stops* is not significant. For the *population* variable this can at least partially be explained by the low standard deviation of the variable (see Table 1). In the case of the *rail stops*, this can be explained by the different characteristics of the local railway system in Hamburg and Hanover (see Section 2.2).

The SLX model shows the same variables to have a significant impact. The number of *jobs*, *shops* and *cultural* facilites has also a significant indirect impact, which is negative for *shops*.

The centrality variable again improves the model fit and is significant and negative for the OLS model. For the SLX model, the introduced variable has no significant impact.

Although the models for Hamburg and Hanover produce different results, most variables have the same sign and a similar size, which shows that demand patterns seem to have comparable structures in both cities.

4.2. Random forest model

The results of the Random forest model for Hamburg and Hanover are shown in Table 4.

Again, we employed one model with and one without centrality variable. The variance that is explained by the model varies from 61.5% to 67%. *Gastronomy* is the most important variable in all four models followed by *jobs*, *culture* and *population* in the case of Hamburg. The low importance of the *population* variable in Hanover supports the (spatial) regression results where the population size had no significant impact. The centrality variable *distance MTS* has the second highest importance if included. *Rail stops* and *shops* have the lowest importance for all four models.

Fig. 3 shows the partial dependence plots for both cities. They show the marginal effect of each variable on the model outcome (number of trips) and indicate their relation (e.g., monotonic or linear).

The overall partial dependencies are similar for Hamburg and Hanover, and except for the distance to the main train station, all variables show a marginally increasing effect. However, some variables show a slightly different pattern in both cities, as for instance *gastronomy* and *distance MTS*. In the case of the main station, the result is not surprising as Fig. 2 showed that the demand in Hanover is concentrated around the city center whereas demand in Hamburg is rather distributed. Interestingly, the variable *shops* also shows an increasing marginal effect

Table 4
Random forest regression results for Hamburg and Hanover with and without centrality variable. The values indicate the variable importance for the overall model.

	Hamburg		Hanover	
		Incl. centrality variable		Incl. centrality variable
Population	11.75	12.78	3.55	3.33
Jobs	12.01	15.95	15.56	12.94
Rail stops	4.39	3.07	3.91	4.78
Dist. MTS		25.19		14.44
Shops	0.40	2.91	3.86	3.47
Culture	9.52	8.85	14.21	13.27
Gastronomy	34.94	28.50	19.32	14.73
Variance explained	61.5%	65.6%	62.8%	67.0%
No. of trees	500	500	500	500

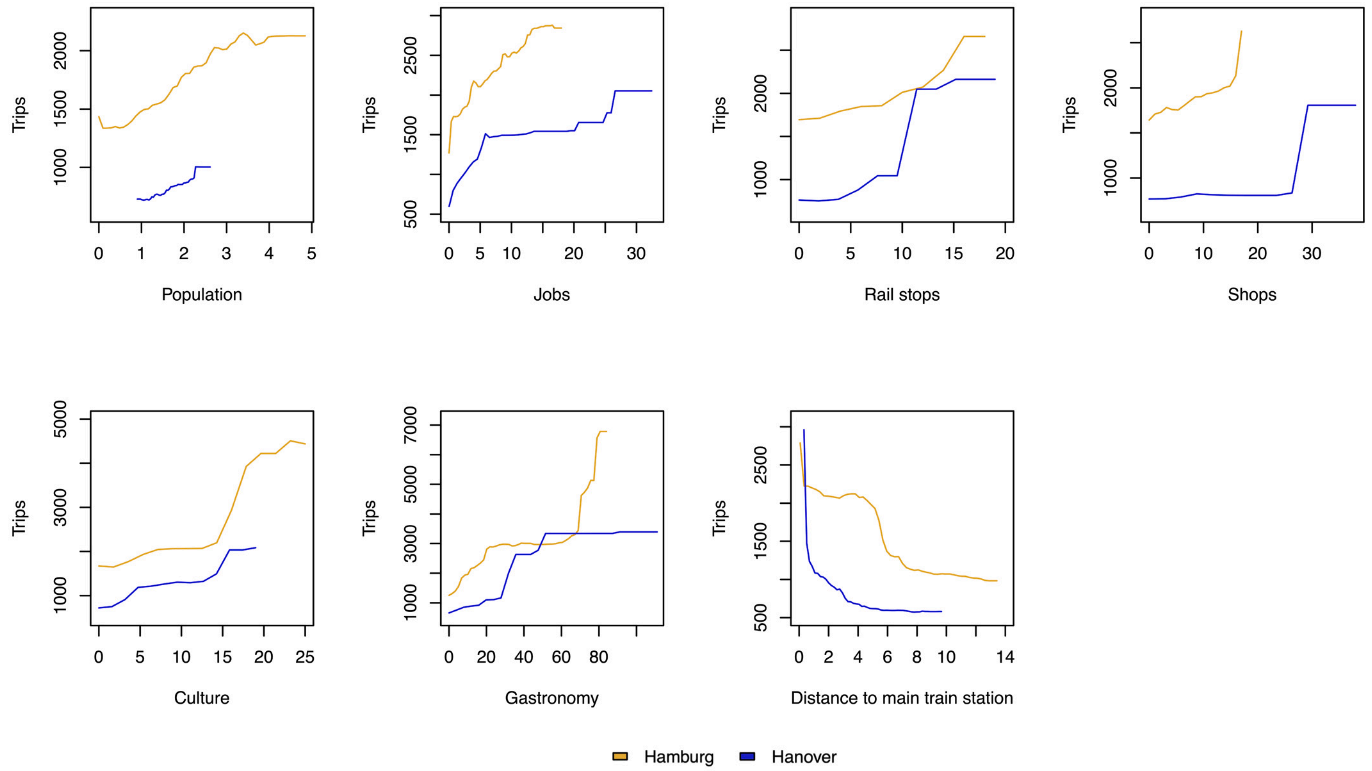

Fig. 3. Partial dependence plots for the Hamburg and Hanover models including the centrality variable.

although its sign in the (spatial) regression models is negative. This could be due to its correlation with the variables *gastronomy* (correlation coefficients in Hamburg/Hanover: 0.52/0.69), *jobs* (correlation coefficients: 0.23/0.52), *culture* (correlation coefficients: 0.30/0.47) and *rail stops* (correlation coefficients: 0.31/0.43). The complete correlation matrices can be found in the appendix.

4.3. Demand prediction

Based on the estimated models for Hamburg, we predicted ride-pooling trips in Hanover. All predictions were made for the specified period from May 2019 to February 2020. We assume that the independent spatial variables in both cities have the same influence on the dependent demand variable. Using the estimated parameters from Hamburg, the independent variables in Hanover are used to make an out-of-sample prediction of ride-pooling trips in Hanover. The results of the prediction are then compared with the real-world observations of MOIA in Hanover.

Table 5 shows the demand prediction errors for predictions with the six estimated Hamburg models. We compare the root mean square error (RMSE), the mean absolute error (MAE) and the overall overestimation of the model with the real data.

All models overestimate the demand in Hanover. The models including the centrality variable *distance MTS* all perform worse than the respective model without centrality variable. Across the model types, the SLX model performs best with a relative overestimation of 18.9% compared to overestimations of 55.0% and 76.3% by the RF model and the OLS model, respectively. The results show that SLX is the best model to control the spatial pattern in the data, indicating that the spillover effects improve the prediction over a simple OLS model. The centrality variable, in contrast, does not control for the spatial pattern uniformly in both cities. The partial dependence plots in Fig. 3 have shown that the distance to the main train station has different impacts in Hamburg and Hanover and is therefore less suitable for prediction.

Fig. 4 shows the absolute and relative spatial distribution of the prediction residuals in Hanover. Red zones indicate that the model underestimated the demand and blue zones indicate that the model overestimated the demand. In most areas, the demand is not over- or underestimated by more than 1000 trips. However, the relative prediction error shows that the model often predicts values that are less than half / more than double of the observed demand. Relative prediction errors seem to be lowest in the city center.

Fig. 5 shows two violin plots with the distribution of absolute and relative prediction errors in Hanover for the specified time period. The horizontal lines show the quartiles and numbers next to the arrows describe the values of extreme outliers. The graphs confirm the aggregated evaluation in Table 5 and the spatial evaluation in Fig. 4 in that the model overestimates demand more often than it underestimates demand. It can be observed that more than 50% of the absolute prediction errors are less than 500 trips off the observed demand. Also,

Table 5
Demand prediction errors of all six estimated Hamburg models applied to Hanover.

	OLS	RF	SLX	Incl. centrality Variable		
				OLS	RF	SLX
RMSE	3641	2941	2863	3963	3616	3262
MAE	663	547	487	796	755	609
Overall relative overestimation of trips	76.3%	55.0%	18.9%	99.4%	89.4%	45.6%
Overall absolute overestimation of trips	229,861	165,633	56,952	299,495	269,542	137,464

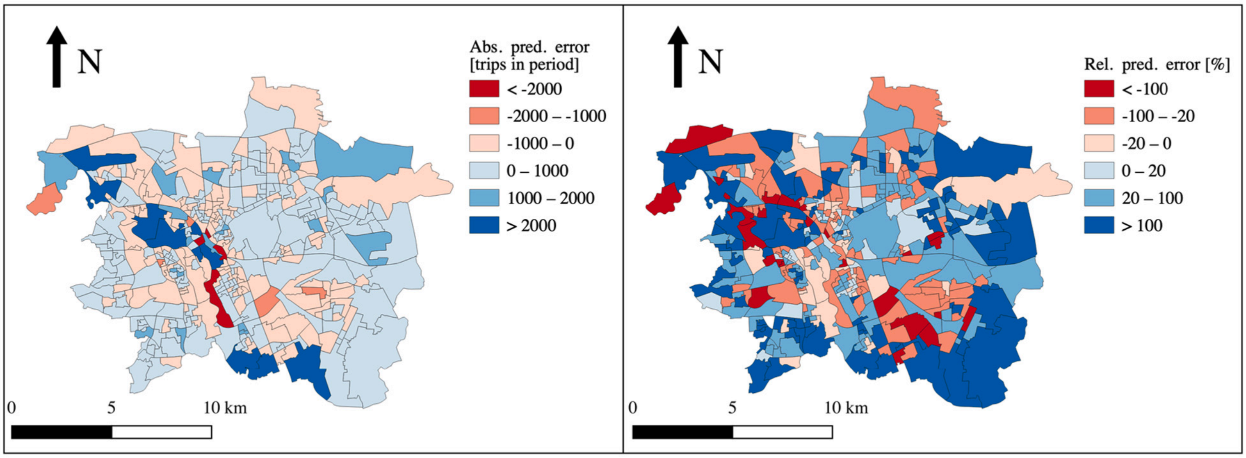

Fig. 4. Spatial distribution of absolute and relative prediction errors per statistical zone in Hanover based on the SLX model without centrality variable. Absolute prediction errors are given for the entire specified period from May 2019 to February 2020. Negative values indicate an underestimation of trip origins. Basic data © Landeshauptstadt Hannover, Geoinformation.

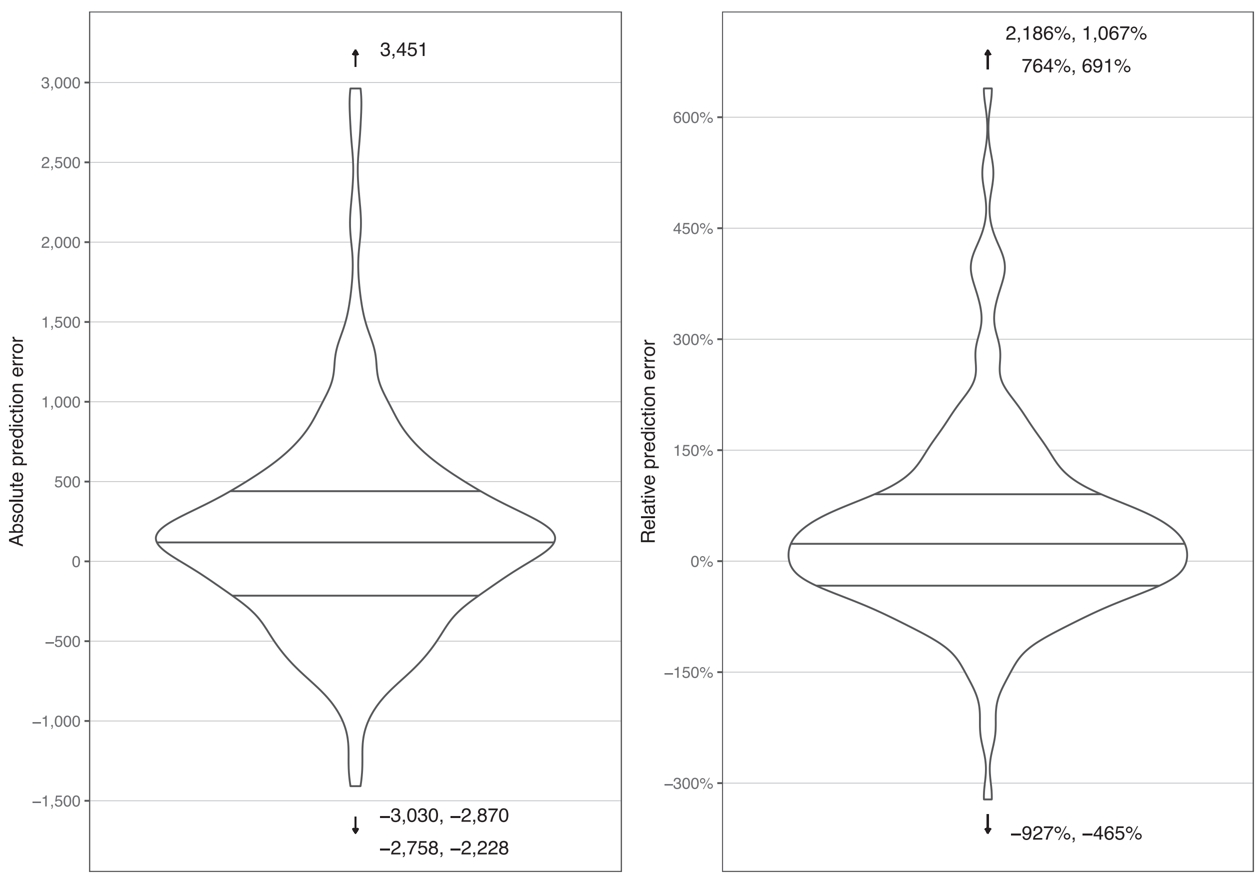

Fig. 5. Violin plots of prediction residuals in Hanover based on the SLX model without centrality variable.

more than 50% of the relative prediction errors are less than 100% above/below the observed demand. However, some extreme outliers are observed for which demand is greatly underestimated or overestimated. These outliers show that while the model provides reasonable predictions for most zones, there are also some zones for which the selected data and model are not at all suitable for predicting demand. With this in mind, a demand forecast for a new city should never be made with the aim of perfect accuracy, but rather to provide an initial understanding of the potential distribution of demand.

5. Discussion and conclusion

In this study, the spatiotemporal demand characteristics of the ride-pooling service MOIA in Hamburg and Hanover were analyzed. After a descriptive analysis of the data, different types of regression models were estimated to understand the spatial characteristics of demand, measured as trip origins, and to identify the transferability of the spatial patterns.

The linear and spatial regression results show significant positive impacts of the parameters *jobs*, *culture* and *gastronomy* and a significant negative impact of *shops* in Hamburg and Hanover. *Inhabitants* and *rail stops* only have a significant positive impact in Hamburg and are not significant in Hanover. The different impacts of rail stops in both cities may be caused by the characteristics of rail-based public transport in both cities, which were explored in Section 2.2. The non-significant regression results of inhabitants in Hanover are likely caused by the properties of the statistical zones in Hanover, which are designed so that all zones contain a similar number of residents with a low standard deviation (see Table 1). As a result, the variable becomes practically a constant and the effect of a variation of the variable cannot be measured. This problem could be solved by using uniform zones. However, some of the spatial data are only available in the specified zones and a transformation of the zonal data would lead to inaccuracies.

Furthermore, the statistical zones in both cities are of different sizes, which can lead to spillover effects of the wrong magnitude being transmitted from one city to the other. These two examples show limitations of spatial transferability and the relevance of validating the input data and comparing their format in both cities.

The positive impact of population, workplaces, (gastronomic and cultural) activity density and transit stops are in line with previous findings from the US and China for ride-hailing demand (Sabouri et al., 2020; Wang and Noland, 2021a; Ghaffar et al., 2020; Yu and Peng, 2019). The significant positive impact of gastronomy and culture confirm the findings of Kostorz et al. (2021) indicating that MOIA is often used for leisure-related trips.

The OLS and the SLX models indicate a negative impact of shops on demand. However, the random forest model shows that demand is higher in areas with more shops, and that the negative impact is caused by the correlation with other variables such as gastronomic and cultural facilities. Consideration of all three models has proven critical to gaining a comprehensive understanding of impacts. For example, the SLX model has shown that while train stations attract ride-pooling demand, the zones around the stations are less in demand. This suggests that ride-pooling is combined with rail-based public transportation, but the areas around subway or train stations are generally less attractive. This result is partially in line with Soria et al. (2020) who also found a positive direct impact of rail access (measured in access time) on ride-pooling usage in Chicago. In contrast to our results, they found a positive spillover effect of rail access on ride-pooling demand. The differing result may be caused by the rail system or the land-use characteristics and would require further investigations.

The inclusion of the centrality variable *distance MTS* contributed to a slight improvement in the OLS and SLX models and the random forest regression showed a high importance of that variable. However, including the variable decreased the predictive accuracy of the models. A potential reason is the differing distribution of demand around the MTS. The partial dependence plots confirm this differing dependence of the centrality variable and trip origins across both cities.

In contrast, the inclusion of a spatial lag of the X variable improved demand prediction substantially. Guidon et al. (2020) had similar findings and did not observe any improvements when centrality variables were included to predict bike-sharing demand in two Swiss cities. In their study, the spatial regression model also has the lowest prediction error, indicating an advantage of such models over OLS or machine learning models to predict spatial mobility demand.

The overestimation of demand in Hanover by at least 18.9%, while not desirable, is still within an acceptable range and much lower than the overestimation that Guidon et al. (2020) indicate for their best performing model. The overestimation can be explained by several known and unknown factors. For example, Hamburg attracts more tourists and business travelers, which the model does not take into account. The airport could be an indication of such additional travel (not just to and from the airport) and needs to be considered in future studies, especially if the area for which demand is predicted includes an airport.

The estimated models may be used for additional out-of-sample demand predictions in new cities. For a disaggregated demand prediction, the descriptive temporal analysis in Fig. 1 provides additional insights into a realistic distribution. The resulting disaggregated demand may then be used for transport simulations to define required fleet sizes serving the expected demand.

Future studies could investigate other centrality variables that might control for the spatial pattern in *both* data sets. More complex spatial models can also be used, but it is questionable whether they are also suitable for predicting demand, since the controlled spatial pattern may vary from city to city.

It can be concluded that although the models presented cannot account for all spatial patterns of demand variation, they are able to adequately predict spatial patterns of demand in new cities. It was shown that it is worthwhile to estimate multiple models, as each model provides additional insight into the data structure. However, the limitations of the models must be taken into account.

The examples of zonal structures (leading to an insignificant impact of population size in Hanover) and different impact of rail stops due to different rail systems in both cities show that one must be very careful when transferring results to new cities. In addition to the zonal structure, the spatial structure of each city is also different and the activity locations have different impacts in other cities. Moreover, the observed POIs are of different size and attraction, which is not taken into account in the model. The estimation measures the average impact of a shop, whether it is a large supermarket or a small kiosk. At city level, these discrepancies within a POI category are likely to be compensated for if the data source is of similar quality, but in parts of a city this can have a significant impact on the results. This must be kept in mind when transferring the results to new cities, especially if they are located outside Germany or even outside Europe.

Two other causes of inaccurate predictions can be different attractiveness of services and different competition in new cities. If the service quality in a new city is not comparable to that in another city, demand is expected to be different. The same applies to competition: if there are already several on-demand services in a city, it is unlikely that a new ride-pooling service will attract a similar number of customers as in a city with little competition. Furthermore, it takes time for demand to adapt to a new service, which is why an established service usually attracts more demand than a recently launched one.

Besides the spatial structures, mobility landscape and service quality, proper planning of an on-demand service also requires consideration of the social, political and economic structures of the cities.

Declaration of Competing Interest

Felix Zwick is currently a PhD candidate at ETH Zurich and studies the impacts of ride-pooling. However, it is also acknowledged that he is employed at the ride-pooling provider MOIA.

Appendix A. Appendix

A.1. Correlation matrices

Table A1
Correlation matrix for (in-) dependent variables and request origins in Hamburg.

	Trip origins	Request origins	Pop.	Jobs	Dist. MTS	Rail stops	Culture	Gastron.	Shops
Trip origins	1.00	0.98	0.26	0.41	−0.45	0.25	0.52	0.76	0.41
Request origins	0.98	1.00	0.24	0.40	−0.40	0.34	0.57	0.78	0.43
Population	0.26	0.24	1.00	−0.17	0.04	−0.03	0.03	0.16	0.29
Jobs	0.41	0.40	−0.17	1.00	−0.35	0.12	0.25	0.46	0.23
Dist. MTS	−0.45	−0.40	0.04	−0.35	1.00	−0.16	−0.26	−0.40	−0.19
Rail stops	0.25	0.34	−0.03	0.12	−0.16	1.00	0.09	0.25	0.31
Culture	0.52	0.57	0.03	0.25	−0.26	0.09	1.00	0.45	0.30
Gastronomy	0.76	0.78	0.16	0.46	−0.40	0.25	0.45	1.00	0.52
Shops	0.41	0.43	0.29	0.23	−0.19	0.31	0.30	0.52	1.00

Table A2
Correlation matrix for (in-) dependent variables and request origins in Hanover.

	Trip origins	Request origins	Pop.	Jobs	Dist. MTS	Rail stops	Culture	Gastron.	Shops
Trip origins	1.00	0.99	0.06	0.67	−0.40	0.51	0.71	0.84	0.55
Population	0.06	0.06	1.00	0.08	0.03	0.10	0.01	0.02	0.09
Jobs	0.67	0.68	0.08	1.00	−0.10	0.52	0.47	0.63	0.52
Dist. MTS	−0.40	−0.35	0.03	−0.10	1.00	−0.05	−0.32	−0.31	−0.19
Rail stops	0.51	0.57	0.10	0.52	−0.05	1.00	0.50	0.48	0.43
Culture	0.71	0.74	0.01	0.47	−0.32	0.50	1.00	0.68	0.47
Gastronomy	0.84	0.84	0.02	0.63	−0.31	0.48	0.68	1.00	0.69
Shops	0.55	0.56	0.09	0.52	−0.19	0.43	0.47	0.69	1.00

References

Aberle, C., 2020. Who benefits from mobility as a service? A GIS-based investigation of the population served by four ride-pooling schemes in Hamburg, Germany. KN J. Cartogr. Geogr. Inform. 70 (1), 25–33.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. U. S. A. 114 (3), 462–467.

Anselin, L., 1988a. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geogr. Anal. 20 (1), 1–17.

Anselin, L., 1988b. Spatial Econometrics: Methods and Models, vol. 4 of Studies in Operational Regional Science. Springer Netherlands, Dordrecht.

Bellon, T., 2021. Lyft to Resume Shared Rides in U.S. for First Time Since Pandemic. https://www.reuters.com/business/autos-transportation/lyft-resume-shared-rides-us-first-time-since-pandemic-2021-07-15/ (Last accessed: 2021-12-22).

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brown, A., 2019. Redefining car access: ride-hail travel and use in Los Angeles. J. Am. Plan. Assoc. 85 (2), 83–95.

Cetin, T., Deakin, E., 2019. Regulation of taxis and the rise of ridesharing. Transp. Policy 76, 149–158.

Chen, X.M., Zahiri, M., Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: an ensemble learning approach. Transp. Res. C 76, 51–70.

Chen, X., Zheng, H., Wang, Z., Chen, X., 2021. Exploring impacts of on-demand ridesplitting on mobility via real-world ridesourcing data and questionnaires. Transportation 48 (4), 1541–1561.

CleverShuttle, 2021. https://www.clevershuttle.de/ (Last accessed: 2021-12-12).

de Souza Silva, L.A., de Andrade, M.O., Alves Maia, M.L., 2018. How does the ride-hailing systems demand affect individual transport regulation? Res. Transp. Econ. 69, 600–606.

Dean, M.D., Kockelman, K.M., 2021. Spatial variation in shared ride-hail trip demand and factors contributing to sharing: lessons from Chicago. J. Transp. Geogr. 91, 102944.

DiDi, 2021. https://www.didiglobal.com/international-business (Last accessed: 2021-12-12).

Engelhardt, R., Dandl, F., Bilali, A., Bogenberger, K., 2019. Quantifying the benefits of autonomous on-demand ride-pooling: a simulation study for Munich, Germany. In: Paper Presented at the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2992–2997.

Ennen, D., Heilker, T., 2020. Ride-Hailing Services in Germany: Potential Impacts on Public Transport, Motorized Traffic, and Social Welfare. working papers 29. Institute of Transport Economics, University of Muenster.

Foljanty, L., 2020. Mapping the Global On-demand Ridepooling Market. https://medium.com/@lukas.foljanty/mapping-the-global-on-demand-ridepooling-market-f8318de1c030.

Gehrke, S.R., Huff, M.P., Reardon, T.G., 2021. Social and trip-level predictors of pooled ride-hailing service adoption in the Greater Boston region. Case Stud. Transp. Policy 9 (3), 1026–1034.

Ghaffar, A., Mitra, S., Hyland, M., 2020. Modeling determinants of ridesourcing usage: a census tract-level analysis of Chicago. Transp. Res. C 119, 102769.

Greenblatt, J.B., Saxena, S., 2015. Autonomous taxis could greatly reduce greenhouse-gas emissions of US light-duty vehicles. Nat. Clim. Chang. 5 (9), 860–863.

Guidon, S., Reck, D.J., Axhausen, K., 2020. Expanding a(n) (electric) bicycle-sharing system to a new city: prediction of demand with spatial regression and random forests. J. Transp. Geogr. 84, 102692.

Halleck Vega, S., Elhorst, J.P., 2015. The slx model. J. Reg. Sci. 55 (3), 339–363.

Hamburg, 2018. Authority for Transport and Mobility Transformation. https://www.hamburg.de/bvm/ (Last accessed: 2021-08-25).

Hanover, 2020. Statistics Office of the State Capital Hanover. https://www.hannover.de/Leben-in-der-Region-Hannover/Politik/Wahlen-Statistik/Statistikstellen-von-Stadt-und-Region/Statistikstelle-der-Landeshauptstadt-Hannover (Last accessed: 2022-02-13).

Henao, A., Marshall, W.E., 2019. The impact of ride-hailing on vehicle miles traveled. Transportation 46 (6), 2173–2194.

Hörl, S., Zwick, F., 2021. Traffic Uncertainty in On-demand High-capacity Ride-Pooling (Preprint).

Hou, Y., Garikapati, V., Weigl, D., Henao, A., Moniot, M., Sperling, J., 2020. Factors influencing willingness to pool in ride-hailing trips. Transp. Res. Rec. 2674 (5), 419–429.

Jing, P., Hu, H., Zhan, F., Chen, Y., Shi, Y., 2020. Agent-based simulation of autonomous vehicles: a systematic literature review. IEEE Access 8, 79089–79103.

Kaddoura, I., Schlenther, T., 2021. The impact of trip density on the fleet size and pooling rate of ride-hailing services: a simulation study. Proc. Comp. Sci. 184, 674–679.

Kagerbauer, M., Kostorz, N., Wilkes, G., Dandl, F., Engelhardt, R., Glöckl, U., Fraedrich, E., Zwick, F., 2021. Ridepooling in der Modellierung des Gesamtverkehrs - Methodenbericht zur MOIA Begleitforschung. Technical Report. Karlsruher Institut für Technologie.

Ke, J., Yang, H., Zheng, Z., 2020. On ride-pooling and traffic congestion. Transp. Res. B Methodol. 142, 213–231.

Knie, A., Ruhrort, L., 2020. Ride-Pooling-Dienste und ihre Bedeutung für den Verkehr. Nachfragemuster und Nutzungsmotive am Beispiel von "CleverShuttle". discussion paper. Wissenschaftszentrum Berlin fÃ¼r Sozialforschung.

Kostorz, N., Fraedrich, E., Kagerbauer, M., 2021. Usage and user characteristics—insights from MOIA, Europe's largest ridepooling service. Sustainability 13 (2), 958.

Kucharski, R., Cats, O., Sienkiewicz, J., 2021. Modelling virus spreading in ride-pooling networks. Sci. Rep. 11 (1), 1–11.

Kuehnel, N., Zwick, F., Hörl, S., 2021. Shifts in Perspective: Operational Aspects in (non-) Autonomous Ride-pooling Simulations (Preprint).

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/ CRC.

Li, W., Pu, Z., Li, Y., Ban, X. Jeff, 2019. Characterization of ridesplitting based on observed data: a case study of Chengdu, China. Transp. Res. C 100, 330–353.

Li, W., Pu, Z., Li, Y., Tu, M., 2021. How does ridesplitting reduce emissions from ridesourcing? A spatiotemporal analysis in Chengdu, China. Transp. Res. D 95, 102885.

Lyft, 2021. https://www.lyft.com/rider/cities (Last accessed: 2021-12-12).

Malik, J., Alemi, F., Circella, G., 2021. Exploring the factors that affect the frequency of use of ridehailing and the adoption of shared ridehailing in California. Transp. Res. Rec. 2675 (5), 120–135.

Martinez, L.M., Viegas, J.M., 2017. Assessing the impacts of deploying a shared self-driving urban mobility system: an agent-based model applied to the city of Lisbon, Portugal. Int. J. Transp. Sci. Technol. 6 (1), 13–27.

Morales Sarriera, J., Escovar Álvarez, G., Blynn, K., Alesbury, A., Scully, T., Zhao, J., 2017. To share or not to share: investigating the social aspects of dynamic ridesharing. Transp. Res. Rec. 2605 (1), 109–117.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1/2), 17–23.

Naumov, S., Keith, D.R., Fine, C.H., 2020. Unintended consequences of automated vehicles and pooling for urban transportation systems. Prod. Oper. Manag. 29 (5), 1354–1371.

Neumann, A., Röder, D., Joubert, J.W., 2015. Toward a simulation of minibuses in South Africa. J. Transp. Land Use 8 (1), 137–154.

OpenStreetMap Contributors, 2019. OpenStreetMap. www.openstreetmap.org (Last accessed: 2021-08-24).

Pernestål, A., Kristoffersson, I., 2019. Effects of driverless vehicles : comparing simulations to get a broader picture. Eur. J. Transp. Infrastruct. Res. 1 (19), 1–23.

Ritzer-Angerer, P., 2021. Sharing Economy trifft ÖPNV — das neue Personenbeförderungsgesetz. Wirtschaftsdienst 101 (10), 789–794.

Ruch, C., Lu, C., Sieber, L., Frazzoli, E., 2020. Quantifying the efficiency of ride sharing. IEEE Trans. Intell. Transp. Syst. 22 (9), 5811–5816.

Sabouri, S., Park, K., Smith, A., Tian, G., Ewing, R., 2020. Exploring the influence of built environment on Uber demand. Transp. Res. Part D: Transp. Environ. 81, 102296.

Soria, J., Stathopoulos, A., 2021. Investigating socio-spatial differences between solo ridehailing and pooled rides in diverse communities. J. Transp. Geogr. 95, 103148.

Soria, J., Chen, Y., Stathopoulos, A., 2020. K-prototypes segmentation analysis on large-scale ridesourcing trip data. Transp. Res. Rec. 2674 (9), 383–394.

Spurlock, C.A., Sears, J., Wong-Parodi, G., Walker, V., Jin, L., Taylor, M., Duvall, A., Gopal, A., Todd, A., 2019. Describing the users: understanding adoption of and interest in shared, electrified, and automated transportation in the San Francisco Bay Area. Transp. Res. Part D: Transp. Environ. 71, 283–301.

Tirachini, A., 2020. Ride-hailing, travel behaviour and sustainable mobility: an international review. Transportation 47 (4), 2011–2047.

Tirachini, A., Gomez-Lobo, A., 2020. Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. Int. J. Sustain. Transport. 14 (3), 187–204.

Uber, 2021. https://www.uber.com/global/en/cities/ (Last accessed: 2021-12-12).

Wang, S., Noland, R.B., 2021a. Variation in ride-hailing trips in Chengdu, China. Transp. Res. Part D: Transp. Environ. 90, 102596.

Wang, S., Noland, R.B., 2021b. What is the elasticity of sharing a ridesourcing trip? Transp. Res. A Policy Pract. 153, 284–305.

Wilkes, G., Engelhardt, R., Kostorz, N., Dandl, F., Zwick, F., Fraedrich, E., Kagerbauer, M., 2021. Assessing the effects of ridepooling - a case study of MOIA in Hamburg. In: Paper Presented at the ITS World Congress 2021, Hamburg.

Young, M., Allen, J., Farber, S., 2020. Measuring when Uber behaves as a substitute or supplement to transit: an examination of travel-time differences in Toronto. J. Transp. Geogr. 82, 102629.

Yu, H., Peng, Z.R., 2019. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. J. Transp. Geogr. 75, 147–163.

Zwick, F., 2017. Analysis and Simulation of Santiago de Chile's Colectivo System. Bachelor's thesis. TU Berlin.

Zwick, F., Axhausen, K.W., 2020a. Analysis of ridepooling strategies with MATSim. In: Paper Presented at the 20th Swiss Transport Research Conference, Ascona.

Zwick, F., Axhausen, K.W., 2020b. Impact of service design on urban ridepooling systems. In: Paper Presented at the 2020 IEEE Intelligent Transportation Systems Conference (ITSC), Rhodes.

Zwick, F., Fraedrich, E., Kostorz, N., Kagerbauer, M., 2020. Ridepooling als Â-ÖPNV--Ergänzung - Der MOIA-Nachtservice während der Corona-Pandemie. Internationales Verkehrswesen 72, 84–88.

Zwick, F., Fraedrich, E., Axhausen, K.W., 2021a. Spatiotemporal Determinants of Ride-pooling Usage in the Course of COVID-19 on the Example of MOIA (Preprint).

Zwick, F., Kuehnel, N., Moeckel, R., Axhausen, K.W., 2021b. Agent-based simulation of city-wide autonomous ride-pooling and the impact on traffic noise. Transp. Res. Part D: Transp. Environ. 90, 102673.

Zwick, F., Kuehnel, N., Moeckel, R., Axhausen, K.W., 2021c. Ride-pooling efficiency in large, medium-sized and small towns - simulation assessment in the Munich metropolitan region. Proc. Comp. Sci. 184, 662–667.