# Expanding a(n) (electric) bicycle-sharing system to a new city

## Prediction of demand with spatial regression and random forests

**Conference Poster**

**Author(s):**
Guidon, Sergio; Reck, Daniel Jan; Axhausen, Kay W.

**Publication date:**
2020-01

**Permanent link:**
https://doi.org/10.3929/ethz-b-000402553

ETH zürich

S. Guidon[1+], D.J. Reck[1], K.W. Axhausen[1]

[1] IVT, ETH Zurich, [+] Corresponding Author; sergio.guidon@ivt.baug.ethz.ch

# Expanding a(n) (Electric) Bicycle-Sharing System to a New City: Prediction of Demand with Spatial Regression and Random Forests

January 2020

## 1 Motivation

The number of bicycle-sharing systems has undergone strong growth in the last two decades. This growth is part of a worldwide trend that began in the 1990s and strongly has accelerated since 2005. Early bicycle-sharing systems have mainly been provided as a public service by cities, but, meanwhile, major international bicycle-sharing companies have emerged that seek to expand their operations to new cities.

Two major strategic questions that arise are which cities should be considered for expansion and the geographical extent of the service area. An important factor to support these decisions is the expected demand for bicycle sharing, as it is directly related to potential revenue.

In this paper, booking data from an electric bicycle-sharing system (Fig. 1 - 3) was used to estimate and assess models for bicycle-sharing demand and to make predictions for an expansion to a new city. Employment, population, bars and restaurants, as well as distance to a central location, were among the most important predictors in terms of variance explained in the same city. However, omitting centrality measures improved predictions for the new city.

## 2 Data Source: Zurich and Berne, Switzerland

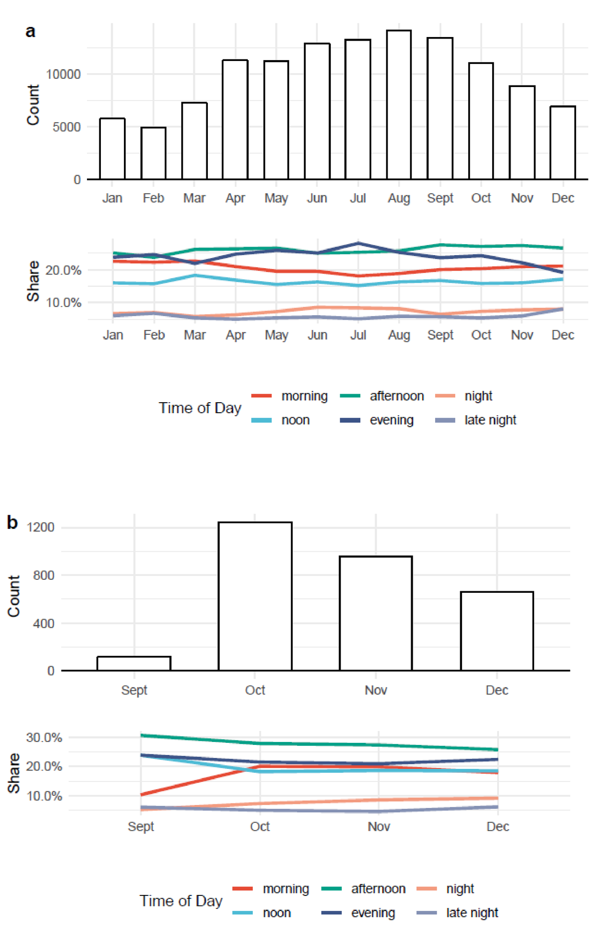

Fig.1: Yearly demand patterns

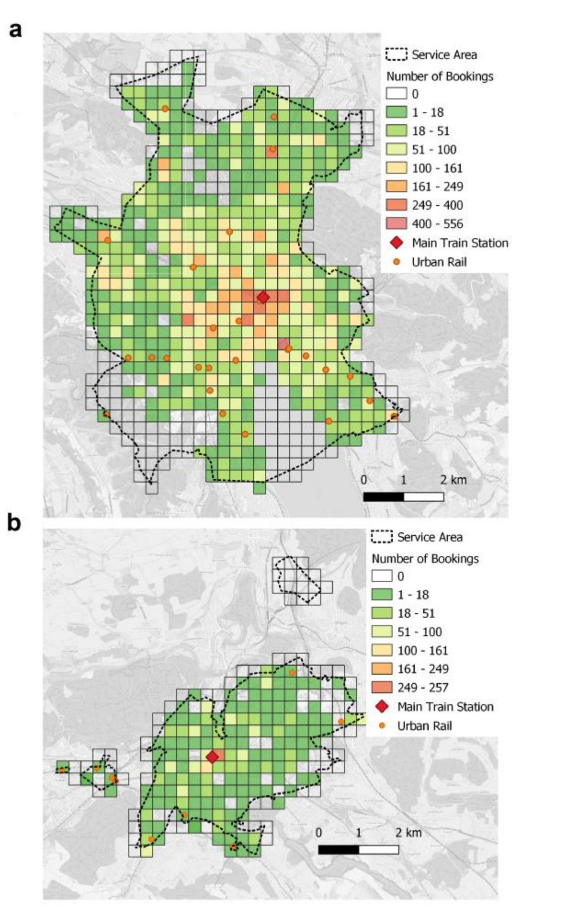

Fig. 2: Spatial Distribution of demand

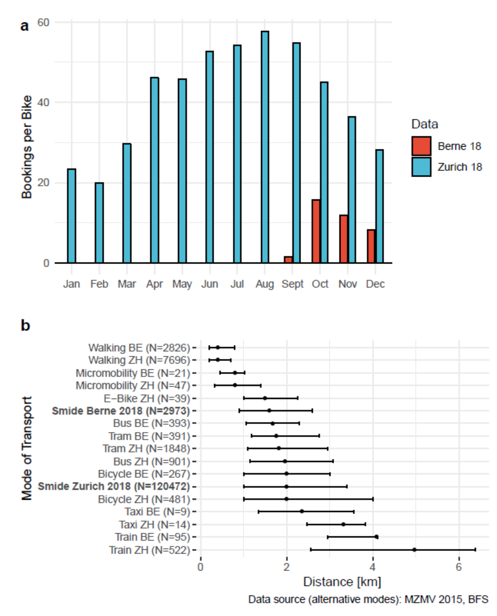

Fig. 3: Data availability and comparison of trip distances

## 3 Model Selection and Specification

Linear and spatial regression models and random forests were estimated for the two Swiss cities of Zurich and Berne (Fig. 4). Models with data from Zurich were then used to predict demand for the city of Berne and the predictions were validated with actual booking data.

The booking data and the spatial variables were aggregated to a 300-meter raster covering the service areas of the two cities. Independent variables included population, work places, public transport availability and measures of centrality.

| Model name | City | Model type | Centrality measures |
|---|---|---|---|
| ZH_LR | Zurich | Linear regression | |
| ZH_SP | Zurich | Spatial regression | |
| BE_LR | Berne | Linear regression | |
| BE_SR | Berne | Spatial regression | |
| ZH_LR_CM | Zurich | Linear regression | √ |
| ZH_SR_CM | Zurich | Spatial regression | √ |
| BE_LR_CM | Berne | Linear regression | √ |
| BE_SR_CM | Berne | Spatial regression | √ |
| ZH_RF | Zurich | Random Forests | |
| ZH_RF_CM | Zurich | Random Forests | √ |
| BE_RF | Berne | Random Forests | |
| BE_RF_CM | Berne | Random Forests | √ |

Fig.4: Estimated demand models

Results from the spatial regression models are shown in Fig. 5 and 6.
Fig. 8 shows the models' fit for the same city.
Fig. 7 and 9 show the prediction errors for the new city.

| | Zurich | | Berne | |
|---|---|---|---|---|
| | ZH_LR_CM<br>Lin. reg<br>Coef./(SE) | ZH_SR_CM<br>SARAR<br>Coef./(SE) | BE_LR_CM<br>Lin. reg<br>Coef./(SE) | BE_SR_CM<br>SARAR<br>Coef./(SE) |
| **number of bookings** | | | | |
| popSize (in thousands) | 18.37*** | 16.64*** | 8.97*** | 9.58*** |
| | (3.49) | (3.10) | (3.24) | (2.88) |
| workPlace (in thousands) | 18.60*** | 15.89*** | 5.78*** | 6.43*** |
| | (1.89) | (1.99) | (1.52) | (1.45) |
| gastronomy (count) | 3.67*** | 2.69*** | 2.73*** | 2.48*** |
| | (0.46) | (0.47) | (0.42) | (0.42) |
| distHB (km) | −12.65*** | −7.95*** | 0.37 | −0.49 |
| | (1.64) | (1.66) | (1.34) | (1.30) |
| boDist (km) | −10.14*** | −8.88*** | −8.70* | −6.73 |
| | (3.20) | (2.42) | (5.00) | (4.20) |
| highPTlevel (dummy) | −4.33 | −5.10* | 2.09 | 2.54 |
| | (3.22) | (2.82) | (3.32) | (2.98) |
| urbanRail200 (dummy) | 38.14*** | 35.96*** | −2.18 | −1.92 |
| | (6.96) | (6.61) | (5.43) | (5.13) |
| HB500 (dummy) | 95.35*** | 77.53*** | 22.28*** | 36.00*** |
| | (15.10) | (14.64) | (7.80) | (8.36) |
| (Intercept) | 62.31 | 38.73*** | 0.87 | 4.27 |
| | (7.71) | (8.06) | (5.33) | (5.51) |
| $\rho$ | – | 0.30*** | – | −0.27 |
| | – | (0.08) | – | (0.21) |
| $\lambda$ | – | −0.35*** | – | −0.23 |
| | – | (0.14) | – | (0.24) |
| $N$ (Number of zones) | 678 | 678 | 248 | 248 |
| AIC | 6918 | 6912 | 2137 | 2130 |
| $R^2_{adj}$ | 0.58 | - | 0.47 | - |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Fig.5: Linear and spatial regression

| | Zurich (RF2_ZH) | | Berne (RF2_BE) | |
|---|---|---|---|---|
| | Variable importance | Rel. variable importance | Var. imp. | Rel. variable importance |
| workPlace | 19.77 | 1.00 | 8.76 | 0.80 |
| distHB | 19.69 | 1.00 | 10.90 | 1.00 |
| gastronomy | 13.89 | 0.70 | 9.38 | 0.86 |
| popSize | 12.76 | 0.65 | 5.95 | 0.55 |
| boDist | 8.33 | 0.42 | 1.38 | 0.13 |
| HB500 | 5.35 | 0.27 | 1.08 | 0.10 |
| highPTLevel | 4.32 | 0.22 | 5.66 | 0.52 |
| urbanRail200 | 2.51 | 0.13 | 0.59 | 0.05 |
| Number of trees | 500 | | 500 | |
| Variance explained | 52.3% | | 26.6% | |

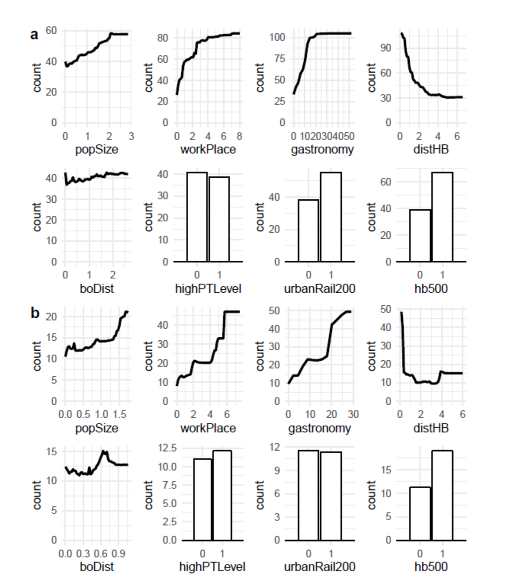

Fig.6: Random forests

## 4 Model Fit and Predictive Performance

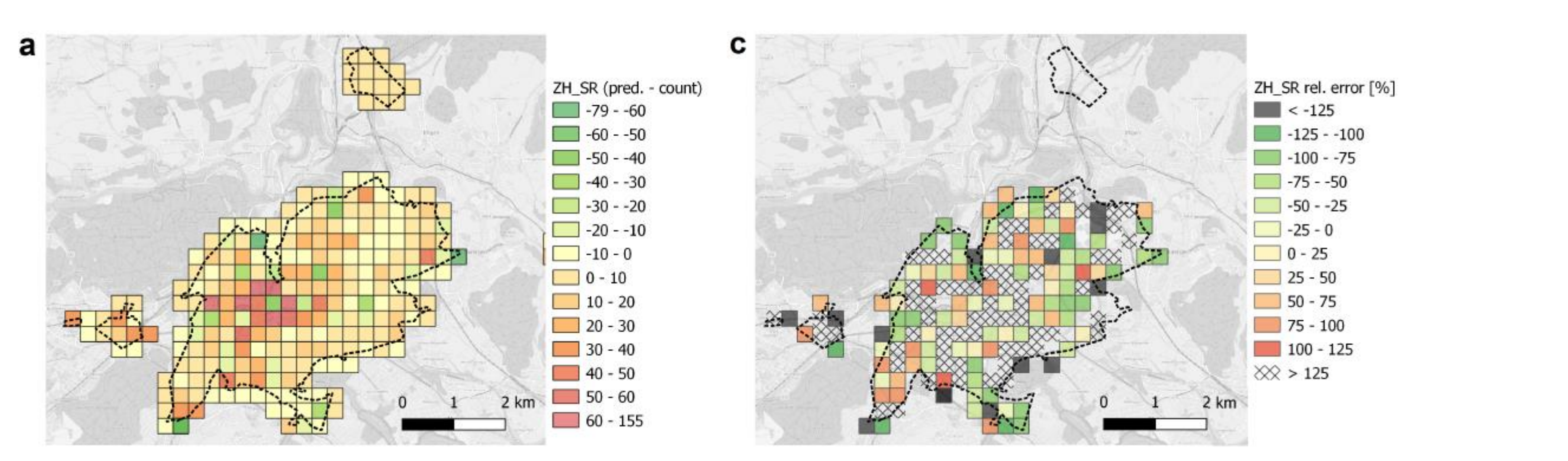

Fig.7: a) absolute prediction error, b) relative error for new city (model ZH_SR, predictions for Berne)

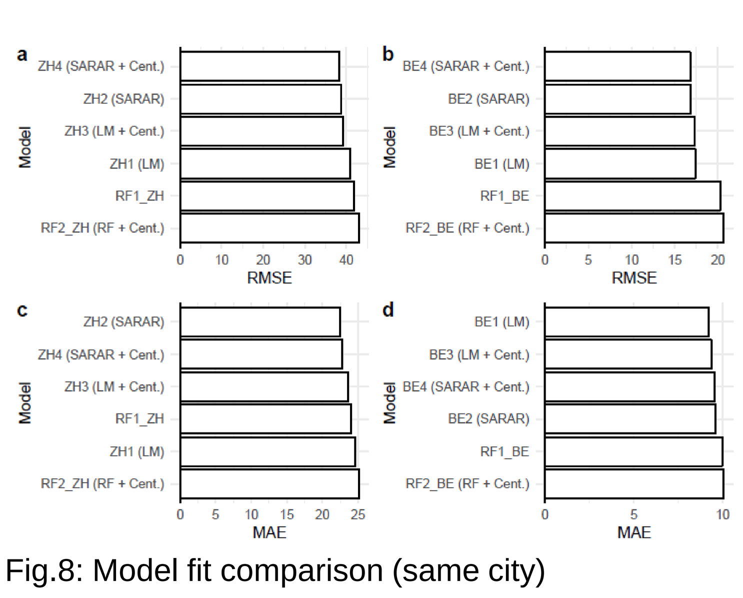

Fig.8: Model fit comparison (same city)

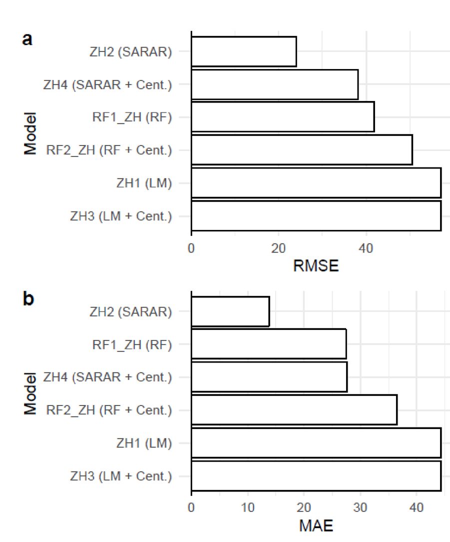

Fig.9: Predictive performance for new city

## 5 Conclusions

Linear and spatial regression models and random forest were estimated to predict demand for an expansion of a bicycle-sharing system to a new city. The most important variables included employment, bars and restaurants and population. Although distance to the main train station and distance to the boundary of the service area (as measures of centrality) improved the model fit, the variables decreased the predictive performance for the new city. Random forests performed worse than spatial regression in this case, although the underlying demand function is most likely not linear. However, spatial regression was able to take into account the spatial dependencies of the data through the neighborhood matrix and was thus supplied with more information than the random forests.

## 6 References

1. Guidon, S., H. Becker, H. Dediu and K.W. Axhausen (2019) Electric Bicycle-Sharing: A New Competitor in the Urban Transportation Market?, Transportation Research Record, 2673 (4), 15-26.
2. Guidon, S., H. Becker and K.W. Axhausen (2019) Avoiding stranded bicycles in free-floating bicycle-sharing systems: Using survival analysis to derive operational rules for rebalancing, IEEE Intelligent Transportation Systems Conference (ITSC 2019), Auckland, New Zealand, October 27-30.

IVT Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

ISTP
Institute of Science, Technology and Policy
Institut für Wissenschaft, Technologie und Politik

+41 44 633 3105
info@ivt.baug.ethz.ch
www.ivt.ethz.ch