GI_Forum 2022, Issue 2
Page: 32 - 46
Full Paper
Corresponding Author:
alexander.rolwes@hs-mainz.de
DOI: 10.1553/giscience2022_02_s32

Analysing and Identifying Geospatial Key Factors in Smart Cities – Model Enhancements in the Use Case of Carpark Occupancy

Alexander Rolwes[1], Pauline Radu[2] and Klaus Böhm[1]

[1]i3mainz, Germany
[2]Mainz University of Applied Sciences, Germany

Abstract

Urban planning benefits significantly from improved knowledge concerning spatiotemporal relationships and patterns in cities based on geospatial factors. In this context, the potential of geodata seems inexhaustible. Applications include limited-service offers like carparks, the occupancy of which is controlled by geospatial factors characterized by their spatiotemporal patterns. This paper proposes an enhanced model for identifying geospatial key factors, tying in with an existing geo-analytics model. Our approach combines real-world empirical data for off-street parking with open-source geodata on points of interest. We formulate stabilization measures in different model-enhancement stages to optimize model reliability and fit, based on analyses of statistical characteristics. Additionally, we consider modifying the choice of geospatial factors in order to reduce multicollinearity. Our results show improved reliability of geo-analytics for the identification of urban spatiotemporal relationships.

Keywords:

geo-analytics, metric of geospatial impact, urban analysis, smart city planning, smart mobility

1 Introduction

One of the major challenges posed by growing cities is the distribution of limited resources such as housing and parking spaces (Boer et al., 2017). Urban mobility suffers from the increasing number of trips made in urban areas by car, which reduce the quality of life in towns and cities (Boer et al., 2017; Giuffrè et al., 2012). Forecasts indicate an increasing burden on transport systems for urban areas. In this context, the importance of flexible solutions for future-oriented urban planning is growing, and smart mobility combines traditional mobility systems with modern communication structures (Zheng et al., 2015). Limited-service offers such as carparks are an integral component of the mobility infrastructure of towns and cities. Thus, optimal use of the available parking spaces is in the interests of urban planners as well as carpark operators and customers.

The power of geospatial data often forms the basis for planning decisions in smart cities. Understanding the relevant geospatial correlations is essential to optimizing mobility processes. Geo-analytics allow us to identify spatiotemporal patterns that influence carpark occupancy. It is also crucial to consider geospatial factors such as nearby food services or shopping facilities that show particular patterns (Cui et al., 2018; Rolwes & Böhm, 2021; Roussel et al., 2022). Geospatial factors trigger carpark occupancy at different times. Therefore, an understanding of this is elementary for the management of urban areas in the future.

This paper expands on and extends the work of Rolwes & Böhm (2021), which focuses on an initial approach to identifying geospatial key factors for urban planning via a metric of geospatial impact. We refer to the work of Rolwes & Böhm as a model for identifying geospatial key factors (MIGKF). This metric describes the impact of geospatial factors; it combines a reachability analysis (see Figure 1) with opening hours and an attractiveness weight of the POI. Statistical results show spatiotemporal relationships between off-street parking and geospatial factors. In addition, the statistical procedures applied allow the model's quality to be assessed. Reliable results increase urban planners' trust in the geospatial relationships. The reliability of the model's results therefore depends significantly on the fulfilment of statistical model prerequisites.

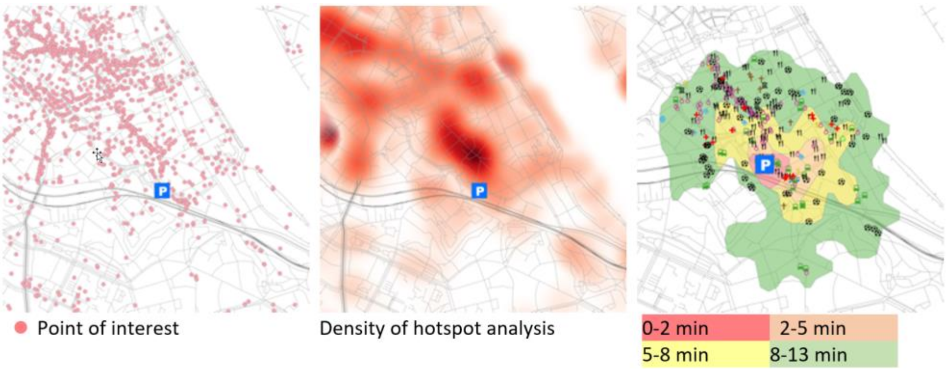

Figure 1: Location-based hotspot and reachability analysis to identify geospatial factors in MIGKF.

We further develop the geo-analytics model to increase the reliability of the results, investigating stabilizing corrections to the existing model based on a comprehensive examination of the model's assumptions. The research question we address is:

In order to identify geospatial key factors in smart cities, how can we stabilize and optimize the existing metric of the geospatial impact in an MIGKF regarding its statistical characteristics?

Smart geospatial data form the basis of future cities (Coors, 2015). By optimizing the statistical parts of geo-analytics, urban planners can gain deeper insights into geospatial key factors and generate benefits for sustainable planning. We use existing real-world empirical data on parking occupancy to examine the research question and examine MIGKF in more detail. This enhanced approach we refer to as Enhanced MIGKF.

2 Related work and challenges

There is ample research on identifying spatiotemporal relationships in smart mobility for urban planning. Common areas of application are e-mobility (Wagner et al., 2013; Wagner et al., 2014), car sharing (Klemmer et al., 2016; Willing et al., 2017), bike sharing (Pelechrinis et al., 2017; Reiss & Bogenberger, 2016; Roussel et al., 2022; Schimohr & Scheiner, 2021; Wang et al., 2021; Wang & Chen, 2020) and parking (Rolwes & Böhm, 2021). Previous studies utilize (historical) POI data, categorize them into geospatial factors, and analyse the spatiotemporal relationships in the use cases in question. These analyses often use conventional statistical models or machine learning algorithms.

Our further development of these earlier studies concerns the optimization of the (geospatial) investigations using a structured metric. By including the opening hours of the POI in its metric, the MIGKF models actual usage patterns in parking behaviour over the day more precisely (Rolwes & Böhm, 2021). In addition, the weighting of the POI in according to the use case reflects the application context's actual characteristics. Carpark data from several years form the basis for developing a metric, which allows well-founded statements on typical parking behaviour. In contrast, other studies look at periods of a few weeks or months, in which deviations from average parking behaviour carry substantial weight.

Considering this background, the extensive dataset of the present study offers great potential for in-depth findings to optimize the MIGKF. As a basis for stabilizing and optimizing the statistical analysis procedures of multiple regression analysis, we consider the work of Field (2017) and Dattalo (2013). In addition, both Field and Dattalo explain the essential model parameters and strategies for testing model assumptions, and highlight further test and correction possibilities to fulfil model requirements.

In summary, these approaches offer a variety of procedures for testing and stabilizing statistical models. Regarding Enhanced MIGKF, we examine the possibility of changing target variables to extend the space-time analysis in the context of the parking behaviour of specific user groups, and focus on optimizing the analysis of geospatial impact factors to increase trust in spatiotemporal relationships.

3 Data analyses and stages of model enhancement

Generally, we apply model exploration techniques on the MIGKF as a base model. By deepening our understanding of underlying model parameters, we identify areas for enhancement and apply four stages of model enhancement (see Figure 2). The results lead to an optimized and stabilized Enhanced MIGKF with improved model performance.

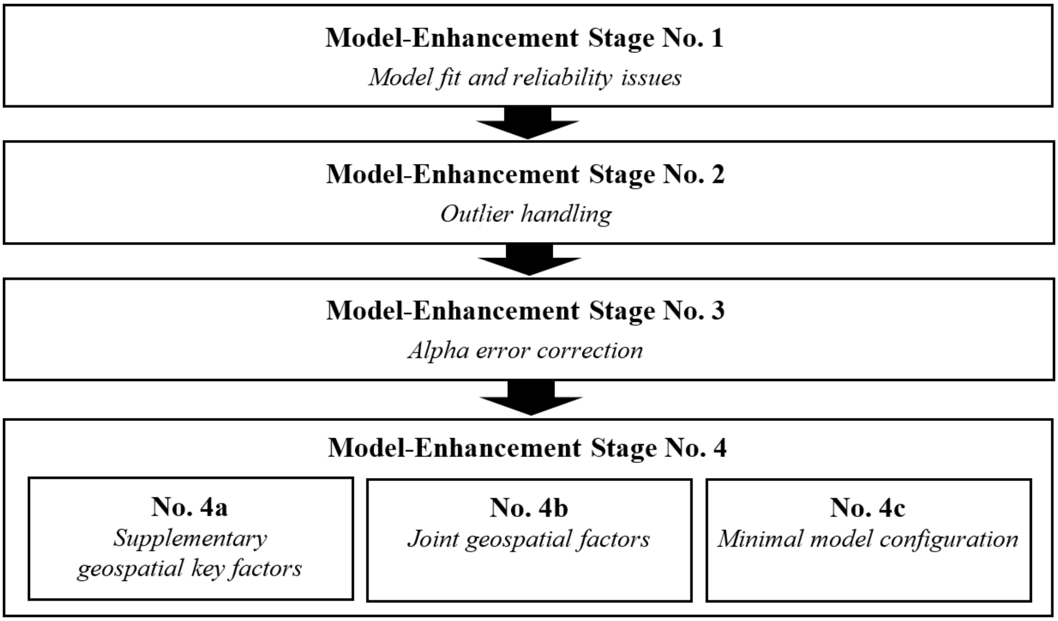

Figure 2: Model-Enhancement Stages for improved model performance in Enhanced MIGKF

3.1 Introduction to the use case

As a starting point, we analyse the existing MIGKF with the aim of gaining a deeper insight into the model. Consequently, the subsequent analyses use data consistent with previous research. Our use case data comprise POI and parking data originating from the city of Mainz, Germany. For this research, the parking occupancy rate of the Kronberger Hof carpark forms a representative example. Its central location close to major shopping streets enables us to analyse the influence of a wide range of nearby POIs. The variety of the POIs provides abundant opportunities for analyses that include numerous possible geospatial key factors on parking behaviour.

The approach allows us to identify distinctive patterns in parking behaviour, in line with the findings of Rolwes & Böhm (2021) in MIGKF. At the same time, we identify areas of possible improvement concerning causality and model power, as well as the quality of results. For example, implicit knowledge of the city led us to expect the model to reveal a higher percentage carpark occupancy during typically busy times, reflected in higher R^2 values. However, as seen for weekday afternoons, for example, this is not always the case (see Table 1). Complementary analyses of typical usage patterns on Google Places confirm this observation. In addition, some findings exhibit unexpected POI patterns. For instance, supplementary analyses imply that food services are an essential geospatial key factor on Saturday evenings for Kronberger Hof carpark. Under the current model, however, these indicators display low scores.

Table 1: User Existing standardized regression results of slot-wise multiple linear regression analysis for Kronberger Hof carpark in MIGKF (Rolwes & Böhm, 2021).

	time of day	services and speciality retail	grocery	health	food services	shopping	adjusted R-squared
Weekday	00:00 - 07:00	0.114 ***	0.000	0.000	0.000	-0.071	0.005
	07:00 - 12:00	0.255 ***	0.034	0.126	0.244 ***	0.171 *	0.639
	12:00 - 18:00	-0.008	0.063	0.240 ***	0.168 ***	-0.133 ***	0.058
	18:00 - 00:00	-0.009	0.020	0.049	0.405 ***	0.132	0.335
Saturday	00:00 - 07:00	0.000	0.000	0.000	-0.109	0.114	0.003
	07:00 - 12:00	0.000	-0.195 ***	0.004	0.747 ***	0.227 *	0.658
	12:00 - 18:00	0.000	-1.478	0.343	0.033	1.616 *	0.141
	18:00 - 00:00	0.008	0.237	0.000	0.155	-0.021	0.005
Sunday	00:00 - 07:00	0.000	0.000	0.028	-0.044 *	0.000	0.002
	07:00 - 12:00	-0.059	0.177	-0.038	0.217 ***	-0.051	0.036
	12:00 - 18:00	0.057	0.059	0.028	-0.145 ***	-0.241	0.015
	18:00 - 00:00	0.158	-0.040	0.049	0.207 ***	0.004	0.090
Significance level at 0.001 (***), 0.01 (**), 0.05 (*), n = 43,815 observations							

We find that the MIGKF requires further investigation. The concerns lie in the limited plausibility of some research outcomes and the lack of model performance in some time slots. In Model-Enhancement Stage No. 1, we call for a detailed assessment of underlying model parameters in order to improve these issues.

3.2 Exploration of statistical characteristics for model choice

By applying model exploration techniques (Dattalo, 2013; Field, 2017), in Model-Enhancement Stage No. 1 we aim for an improved understanding of MIGKF. With this in mind, we utilize statistical tests alongside visual analysis to find causes for the weaknesses in model performance and plausibility outlined above. The model's accuracy relies on the assumption that specific model characteristics apply. For example, in adopting a linear relationship between geospatial key factors and parking behaviour, we assume a linear regression model as a basis, as found in other use cases (Klemmer et al., 2016; Willing et al., 2017). A careful analysis of the model's traits promises a deepened understanding of the model alongside more reliable results.

Before focusing on questions of model fit and reliability, we first validate the overall model choice by considering basic model properties. Our analysis uses parking occupancy and POI categories as continuous variables. Since a linear regression model may be applicable for continuous variables, essential characteristics justify the choice of base model as a starting point. To further validate the choice of a linear model, we investigate characteristics of additivity and linearity. Accordingly, we assess P-P plots supplemented by residual histograms. For named analyses, we consider each regression model for each time slot in order to obtain a general picture of the characteristic in question. In our particular case study, eight out of twelve time slots exhibited minor weaknesses in the characteristics being examined, which manifested as discrete deviations from symmetrical patterns. Furthermore, we noted several significant deviations, far outside the mean, indicating outlier influence as a possible cause. Importantly, we did not observe non-linear patterns, which would have indicated the need for a non-linear base model. As part of an analysis of model fit, we address outliers further in the next section (3.3).

Dattalo (2013) recommends evaluating autocorrelation based on statistical assumption. This requires the time-series data to be close to random. In our case study, the Durbin-Watson test statistic reveals high autocorrelation, with scores below 1 in every time slot and indicators displaying values between .112 and .694. Generally, autocorrelation is typical for time-series data. Re-occurring temporal patterns inherent in parking data result in a shifted offset in correlation as a measure of time. Distinctive temporal patterns become apparent over periods of a day or a week. Our analyses show that carparks have generally higher occupancy during the day than at night. However, each discrete occupancy value displays little change from one hourly score to another. This behaviour is part of the data, as observed in similar use cases (Klemmer et al., 2016; Schimohr & Scheiner, 2021). As a consequence of autocorrelation, temporal dependency can cause elevated alpha errors, hindering an accurate interpretation of the model. Autocorrelation remains a limiting factor, since it cannot be eliminated (Dattalo, 2013). We explore appropriate correcting measures for elevated alpha errors in Section 3.3.

Although at the end of Model-Enhancement Stage No. 1 linear regression remains a suitable model, we should consider limiting factors like elevated alpha errors and outlier influence when focusing on model fit.

3.3 Outlier handling: investigation of factors limiting model performance

Building on the factors hindering model performance discovered in the previous stage, in Model-Enhancement Stage No. 2 we focus on outlier handling. Specifically, we investigate outliers for their characteristics, causes and possible solutions, in order to identify areas of overall model optimization. With the objective of increased reliability of results paired with an improved model fit, we further develop the Enhanced MIGKF. To quantify the impact on the model caused by outliers, we focus on two model characteristics: normality in residuals and heteroscedasticity.

Ideally, residuals follow a normal distribution. Accordingly, most of the observations account for low residuals, indicating that the model generally matches the data well. However, visual examination of histograms recording residual distribution shows mixed results. Half of the time slots exhibit skewed distributions, with outliers appearing as distinctive deviations. In addition, we consider characteristic values of the residual distribution. These indicate how far the distribution skews from average, which may correspondingly indicate the scope of outlier impact. Six time slots show weaknesses in this instance, with kurtosis values exceeding ± 1. Regarding skew, two time slots display values exceeding ± 1 (Field, 2017).

Greene (1993) and Verbeek (2017) confirm extreme values as common causes of heteroscedasticity. We detect an increase in variance for predicted values rising in three out of twelve time slots. Residuals demonstrate a fanning shape, resulting in fluctuating standard deviations, thus limiting the degree of reliability we achieve. We observe standard deviations of selected observations: up to six standard deviations falling outside the norm are paired with extremely low or high residuals, indicating corresponding atypical observations. However, all time slots show outlier influence. As a result, the presence of outliers in large numbers limits model reliability, because including atypical observations hinders the precise modelling of geospatial factors.

To distinguish different causes for atypical occupancy rates, we identify and group outliers according to possible causes, by considering standardized deleted residuals (SDR). For this purpose, we compare the SDR values of more than 3 with supplementary data provided by the carpark owner (Field, 2017; Huber, 2004; Velleman & Welsch, 1981). In the process, we match most outliers to dates of on-site construction, events and public holidays as possible external factors causing atypical parking behaviour. Consequently, these values do not closely represent the data we intend to examine, and temporal POI characteristics cannot adequately model parking behaviour for these timeframes. In such cases, Field (2017) recommends the removal of outliers.

Thus, an updated model excludes outliers for enhanced model composition. We recognize that excluding selected values bears the risk of introducing bias. Likewise, we recognize that dropping observations results in a reduced sample size. However, since each time slot comprises at least 1,000 data rows, we can exclude a large number of outliers without adversely affecting the minimum sample size (Harrell et al., 1996). We apply an iterative process to eliminate the outliers based on their corresponding SDR values. This case-by-case consideration of observations excludes only those values that risk distorting results, a process which allows for a maximized sample size while including a wide range of observations. However, we acknowledge that this leads to many minor outliers remaining in the model.

After performing the elimination processes in Model-Enhancement Stage No. 2, we note enhanced model parameters overall. Values for skew and kurtosis now show patterns closer to normality, and four additional time slots display a normal distribution. Most importantly, homoscedasticity improves massively. Before eliminating outliers, three slots showed homoscedasticity. However, this statistic jumps to ten, exhibiting homogeneous patterns in seven additional time slots. Finally, linearity also shows improvements in seven slots. Overall, Akaike information criterion (AIC) scores become lower in every time slot, while R^2 scores improve in eight. These metrics indicate that the updated model is able to describe a higher percentage of parking behaviour. The model shows improved reliability and goodness of fit.

3.4 Alpha errors: effects of correcting measures on significance levels

After removing outliers, some unexpected results persist. For example, on weekday nights *health* appears as a significant geospatial factor ($b = .119$; $p = .012$) now. However, as most healthcare POIs close in the late afternoon, we would expect a lower significance. We attribute this lack of model plausibility to elevated risks of alpha errors, as described in the literature (Dattalo, 2013; Field, 2017). Alpha errors cause false identification of geospatial key factors as significant, provoking non-plausible results. Consequently, as part of Model-Enhancement Stage No. 3, we consider the application of alpha error correction to validate results and improve overall model reliability.

We acknowledge that an adjustment of alpha levels may decrease statistical power (Moran, 2003; Nakagawa, 2004). In addition, we note that the necessity of alpha error correction measures and their actual implementation is the subject of controversy (Armstrong, 2014; Cabin & Mitchell, 2000; Perneger, 1998). However, we justify correcting measures because of the high autocorrelation present in the use case. Applying such measures leads to increased reliability of relevant patterns: afterwards, for example, *health* shows lower significance on weekday nights ($b = .119$; $p = .048$). Overall, 16 different influences remain significant ($p < .05$), with 12 time slots exhibiting strong significance. Thus, in Model-Enhancement Stage No. 3, we preserve major patterns of the previous MIGKF while improving plausibility.

3.5 Exploration of alternatives in model composition

Rolwes & Böhm (2021) state that multicollinearity in independent variables further limits the reliability of results in the MIGKF. In their use case, the limited separation of geospatial impact factors blurs the information expressed by each variable. As a result, geospatial impact factors express a reduced explanatory power individually (Field, 2017). In a further step in creating the Enhanced-MIGKF, we investigate this remaining issue. We differentiate measures to reduce multicollinearity as Model-Enhancement Stages No. 4a to 4c. As a basis for enhancement measures, we propose a separate evaluation of the detailed characteristics of each geospatial impact in order to assess the extent of multicollinearity present in the model. We examine variance inflation factors (VIF) and pairwise correlations for an overview. Results confirm high multicollinearity in ten out of twelve time slots, based on a cut-off for VIF values set at ten (Bowerman & O'Connell, 1990; Ziegel & Myers, 1991). We suspect the causes of overlapping POI information are rooted in geospatial and temporal correlation. Regarding spatiotemporal correlations, similarities in opening hours have a greater influence in the morning and decrease later in the day. Thus, nuances in temporal patterns separate influences of different categories: we observe lower multicollinearity for variables displaying distinctive opening hours. For example, *food services* are open until late at night, whereas *health* POIs such as doctors' surgeries close much earlier. These stand out due to their distinctive temporal patterns.

The geospatial patterns of the categories modelled also share similarities. For instance, retail stores and service places cluster around main roads and major shopping streets, thus providing many POIs within similar walking distances. As spatial interdependence constitutes a fundamental characteristic of geo-data in cities, we can hardly avoid multicollinearity. Bendler & Ratku (2014) note a high degree of linear interdependence for a similar use case. Ultimately, model characteristics inhibit the complete elimination of multicollinearity, while measures minimizing their effect remain viable. With this in mind, we proceed to Model-Enhancement Stages 4a to 4c.

Beginning with Model-Enhancement Stage No. 4a, using 5 criteria we explore the inclusion of a supplementary geospatial key factor in the MIGKF and its potential to improve the model. For this purpose, we add all available geospatial factors into the MIGKF and examine the results. A broader model that includes several additional geospatial factors demonstrates rising VIF scores combined with worse model characteristics overall. Although additional POI information accounts for negligible further improvement in model fit, the investigation of all POI data deepens our understanding of variable characteristics. Based on this, we ask which criteria to apply when selecting suitable supplementary variables. Referencing the MIGKF, Rolwes & Böhm (2021) choose POI categories based on how representative the underlying opening hours are. They prioritize well-recorded categories with a high percentage of POIs for which the opening hours are known. Accordingly, we prefer to treat the geospatial impacts of geospatial factors for which we have plausible or known opening hours. In this manner, influence scores represent actual category characteristics more accurately.

We consider both the levels of completeness and the plausibility of temporal patterns when evaluating the addition of geospatial impacts. To further prioritize plausibility, for each geospatial factor we factor in the types and the number of POIs that are within walking distance. Finally, we minimize the correlation between the categories.

Overall, we observe that geospatial factors fit the proposed criteria to varying extents. Examples of good fit include POIs in the category *food services*. Their distinctive opening hours make for a less correlated category. In addition, the actual POIs surrounding the carpark comprise many restaurants, cafés and bars. We conclude that *food services* is a distinctive variable with uniform POI characteristics.

After careful consideration, we chose *public sector* as an additional variable to test for in Stage No. 4a. This category has high potential usage by visitors and employees of the many government agencies nearby. Additionally, many POIs for the public sector have distinctive opening hours, being closed to the public in the afternoon. Compared to other categories' opening hours, they display lower VIF scores overall. We test the perceived importance of the variable *public sector* by including the relevant POIs in the model.

The five geospatial criteria in the MIGKF listed above provide the basis for evaluating the impact of other geospatial impacts. Findings, however, fail to show any improvement in model quality. We confirm this by comparing the AIC of this model to that of the previous version (i.e. the version constructed in Section 3.4). Five time slots show no change in AIC, while four worsen. Just three slots indicate improved AIC values, and only one displays an improved R^2 score.

To sum up the results of Model-Enhancement Stage No. 4a, we conclude that an increase in variables fails to improve the model's explanatory power, despite careful selection processes.

In Model-Enhancement Stage No. 4b, we evaluate the potential benefits of reducing the number of geospatial factors. We combine variables as joint geospatial key factors to reduce multicollinearity (Field, 2017; Frost, 2020). By merging categories manually, we factor in the five criteria discussed above. Moreover, combining categories that share similarities in temporal and geospatial patterns preserves interpretability. In this case, we merge *shopping* and *services and speciality retail*. Our choice is justified because of the blurred category border (some shops may also offer speciality retail). As a result, we note a substantial overlap of the geospatial patterns of the two categories, which show a similar clustering along the main shopping streets. The POIs share further similarities in terms of opening hours and potential frequency of use.

After including *shopping and services* as a single factor in place of its two components separately, the combined geospatial factor displays patterns common to both initial factors. We notice an overall improvement in VIF values, leaving just eight slots with suboptimal correlations.

Before combining categories, results show just one time slot with VIF values below ten. By combining categories, we achieve a more concise representation of model variables. At the same time, however, we observe worse AIC scores. To sum up the results of Model-Enhancement Stage No. 4b: the combination of geospatial factors represents a considerable improvement in model reliability but at the cost of a weaker model fit.

Building on the changes to the model described above, we explore a minimal model configuration as the final stage (Model-Enhancement Stage No. 4c), which involves a final selection process in which we gauge the extent of the model performance sustained while minimizing correlation. Field (2017) notes that all variables demonstrate high correlations in the face of high multicollinearity. Excluding one variable may therefore lower multicollinearity, but this rarely rules it out completely (Frost, 2020; Zellner et al., 2001). Nonetheless, we test excluding one variable, leaving a selection of the most distinctive POI categories. In doing this, we were aiming for lower VIF scores while maintaining model performance and decided to exclude *grocery* as a category. Earlier analyses of parking according to user groups revealed that customers are more likely to use supermarket and other store carparks than municipal carparks.

If a carpark had only a small number of POIs in its immediate vicinity, we decided to categorize the carpark itself as a secondary destination for customers (i.e., we assumed that customers must be using the facility to access other POIs further away).

After excluding just one geospatial factor, half of the time slots exhibit improved AIC and R^2 values, while the other half display worse results. Meanwhile, VIF scores sink lower. Half of the time slots improve significantly, finally showing acceptable VIF scores and thus lower multicollinearity. At the same time, other model parameters for the most part remain constant. We notice that the slimmer model used in Stage 4c, which uses fewer variables, shows an improved interpretation of the underlying effects. Lower correlation scores lead to a more reliable and precise interpretation of the use case. On the downside, we observe a loss in explanatory power when excluding variables beyond the five mentioned in MIGKF.

In summary, to limit multicollinearity we split the exploration of variations in model composition into three enhancement stages. First, Model-Enhancement Stage No. 4a examines the inclusion of a supplementary geospatial impact. Next, in Stage No. 4b, we combine geospatial factors which share main characteristics. Finally, we reach a minimal model configuration in Stage No. 4c. Inaccuracies in our data for opening hours lead to a loss in precision, however. Other limiting factors include the similarities mentioned above in temporal and geospatial patterns. Despite these limitations, we have been able to optimize Enhanced MIGKF and to attribute the information derived from the model to POI categories more clearly.

4 Results and discussion

In conclusion, we stabilize the model by exploring the different Model-Enhancement Stages. Identifying and excluding outliers based on their characteristics and significance for the use case at hand leads to improved goodness of fit. Additionally, improved residual characteristics indicate a more reliable model. To further enhance reliability, we apply alpha error correction. Lastly, combining geospatial factors minimizes multicollinearity. As a result, we are able to model the geospatial impact more clearly, reducing blur.

Here, we apply three criteria for different model variations to limit multicollinearity. We consider the completeness of information regarding opening hours, the composition of POIs in this category, and correlations. Variants for which fewer variables were included showed lower VIF. However, as variations in model composition reach varying degrees in the goodness of fit, there is no ultimate or best model. We describe the resulting trade-off observed as follows.

In reduced models, non-optimal correlation scores remain. Further, we observe lower model fit according to R^2 and AIC measures. Here, issues arise when a lower number of variables engenders misleading results, possibly implying connections not valid for a more generalized context. This drawback hinders a general statement concerning the underlying relationships. In contrast, we observe high multicollinearity in a broader model. As seen in Section 3.5, achieving the best choice of geospatial factors is critical in under- and overfitting. An enhanced (i.e. minimal) choice increases the number of significant time slots from 16 to 18 ($p < .05$), and the number of time slots with high significance increases from 12 to 15 ($p < .001$). In addition, we observe improved R^2 values in all but two slots, indicating improved model performance. Finally, we characterize the enhanced model for identifying geospatial key factors (see Table 2).

Table 2 clearly shows that the geospatial factor *food services* indicates the highest positive effect on parking occupancy in the middle of the day on Saturdays ($b = .767$; $p < .001$). Consequently, this category represents a geospatial key factor. We infer that restaurants and other destinations centred on *food services* account for a significant portion of customers on Saturdays. For example, on weekday nights R^2 values change from .335 to .449. Other changes range from .658 to .759 around midday on Saturdays and from .141 to .298 on Saturday afternoons. Additionally, before enhancing measures are applied, *health* displays a single significant time slot on weekday afternoons (see Table 1).

Table 1: Standardized regression results of slot-wise multiple linear regression analysis of the Kronberger Hof carpark in Enhanced MIGKF.

	time of day	shopping and services		food services		health		adjusted R-squared
Weekday	00:00 - 07:00	0.093		-0.044		0.000		0.002
	07:00 - 12:00	0.500	***	0.267	***	0.167	***	0.814
	12:00 - 18:00	-0.165	***	0.253	***	0.338	***	0.111
	18:00 - 00:00	0.146	**	0.484	***	0.061		0.449
Saturday	00:00 - 07:00	0.000		-0.002		0.000		0.001
	07:00 - 12:00	0.250	***	0.767	***	-0.166	***	0.759
	12:00 - 18:00	0.575	***	0.002		-0.028		0.298
	18:00 - 00:00	-0.522		0.411	***	0.523		0.181
Sunday	00:00 - 07:00	0.000		-0.054		0.031		0.004
	07:00 - 12:00	0.165		0.064		-0.008		0.035
	12:00 - 18:00	-0.338	*	-0.182	***	0.346	*	0.037
	18:00 - 00:00	0.172	***	0.186	***	0.065		0.103
Significance level at 0.001 (***), 0.01 (**), 0.05 (*), n = 34,846 observations								

After enhancement, four significant time slots become apparent ($p < .05$). However, we acknowledge that non-plausible results persist to some extent, notably two significant time slots for *health* POIs on Sundays when none of them are open. We see this as an indication of spurious correlations. In addition, some patterns still pair with low R^2 values, although the remaining significant time slots display plausible patterns overall. For example, the highest coefficient for *food services* appears on Saturday noon ($b = .767$; $p < .001$), pairing with an R^2 value of .759. Other plausible examples include Saturday evenings ($b = .411$; $p < .001$).

In conclusion, *shopping and services*, *health* and *food services* greatly influence occupancy rates for this carpark. Applying the statistical adaptations that we have described, our Enhanced MIGKF displays these patterns more clearly and reliably: reliable geospatial key factors influencing parking behaviour emerge using Enhanced MIGKF. We recommend the approach for application scenarios with a linear base model and comparable model composition. The results can be used by urban planners for better and more reliable identification of geospatial relationships.

5 Conclusion and future work

Sustainable resource management remains an essential issue in growing cities. Resource management and the lack of sustainably designed areas present urban planners with challenges in the design of transport space, including parking. Urban planners and domain experts therefore seek to understand geospatial key factors determining parking behaviour. For improved decision-making in carpark management, we propose an enhanced model for the identification of geospatial key factors which includes adjusted spatial core data. We stabilize existing approaches by applying different correcting measures and merging distinctive geospatial key factors. Overall, the Enhanced MIGKF offers tools to improve the reliability of geo-analytics and the utilization of underlying geodata. Our approach improves knowledge of urban geospatial key factors that affect parking behaviour. Using our approach, urban planners can better understand spatiotemporal relationships and mobility dynamics, i.e. why a carpark has high or low occupancy at different times. Our analysis can add value to decisions around innercity changes, as in location optimization or site planning. Decisions such as where to build new carparks or which can be closed can be supported in part by information on the spatial environment. Taking into consideration the opening hours of customers' possible destinations remains an advantage in optimized geo-analytics models, but the model is not limited to carparks: it could be applied, for example, to bike-sharing services or charging stations for electric vehicles.

The model versions with differing compositions achieve improvement in multicollinearity and goodness of fit, if to different degrees. We therefore suggest adapting geospatial key factors concerning data completeness to improve the model's temporal patterns. Building on this, future research might conclude that other combinations of POI categories or merged POI categories create a more fitting model. Integrating statistical approaches benefits a wide range of use cases. To this end, we recommend ongoing dialogue with experts in the field. In general, we emphasize the need for geo-visualization to support urban planning. Communicating results with visualizations increases their accessibility to experts, thereby facilitating their practical application based on a deepened and enhanced understanding of geospatial key factors. Furthermore, we recommend an exploratory approach by offering domain experts the possibility of changing input parameters, for example testing a different combination of geospatial factors or changing model variables. As geo-analytics become more accessible, trust in practical applications resulting from its findings will increase, while attention to feedback from experts will allow further research and development. Overall, our results contribute considerably to an improved decision-making process in smart city design and resource management.

Acknowledgements

The research for this paper is part of the project 'BAM – Big-data-analytics in environmental and structural monitoring' at Mainz University of Applied Sciences, funded by the Carl Zeiss Foundation. The authors would like to thank the local parking company for its excellent cooperation and provision of data. They are also grateful for the proofreading of their article. This study is part of ongoing PhD research at RheinMain University of Applied Sciences and Mainz University of Applied Sciences.

References

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

Bendler, J., & Ratku, A. (2014). Crime Mapping through Geo-Spatial Social, *2014*.

Boer, R., Paundra, J., & Brandt, T. (2017). Know Before You Go: Predicting Parking Space Occupancy by Exploiting Publicly Accessible Data.

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear Statistical Models: An Applied Approach*. International Thomson Publishing.

Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or Not to Bonferroni: When and How Are the Questions. *Bulletin of the Ecological Society of America*, *81*.

Coors, V. (2015). Ohne smarte Geodaten keine smarten Städte. *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement*(4/2015), 244–248. https://doi.org/10.12902/zfv-0076-2015

Cui, Y., Meng, C., He, Q., & Gao, J. (2018). Forecasting current and next trip purpose with social media data and Google Places. *Transportation Research Part C: Emerging Technologies*, *97*, 159–174. https://doi.org/10.1016/j.trc.2018.10.017

Dattalo, P. (2013). *Analysis of Multiple Dependent Variables: Pocket Guide to Social Work Research Methods*. Oxford University Press.

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th edition). SAGE Publications.

Frost, J. (2020). *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*. Jim Publishing.

Giuffrè, T., Siniscalchi, S. M., & Tesoriere, G. (2012). A Novel Architecture of Parking Management for Smart Cities. *Procedia - Social and Behavioral Sciences*, *53*, 16–28. https://doi.org/10.1016/j.sbspro.2012.09.856

Greene, W. H. (1993). *Econometric analysis* (2. ed.). Prentice Hall.

Harrell, F., Lee, K. l., & Mark, D. B. (1996). Tutorial in Biostatistics: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Tutorial in Biostatistics*, *14*.

Huber, P. J. (2004). *Robust Statistics*. John Wiley & Sons.

Klemmer, K., Willing, C., Wagner, S., & Brandt, T. (2016). Explaining Spatio-Temporal Dynamics in Carsharing: A Case Study of Amsterdam. *Americas Confrence on Information Systems.* Advance online publication. https://doi.org/10.13140/RG.2.1.3510.8721

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, *100*(2), 403–405. https://doi.org/10.1034/j.1600-0706.2003.12010.x

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology*, *15*(6), 1044–1045. https://doi.org/10.1093/beheco/arh107

Pelechrinis, K., Zacharias, C., Kokkodis, M., & Lappas, T. (2017). Economic impact and policy implications from urban shared transportation: The case of Pittsburgh's shared bike system. *PLOS ONE*. Advance online publication. https://doi.org/10.1371/journal.pone.0184092

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, *316*(7139), 1236–1238. https://doi.org/10.1136/bmj.316.7139.1236

Reiss, S., & Bogenberger, K. (2016). Validation of a Relocation Strategy for Munich's Bike Sharing System. *Transportation Research Procedia*, *19*, 341–349. https://doi.org/10.1016/j.trpro.2016.12.093

Rolwes, A., & Böhm, K. (2021). Analysis and evaluation of geospatial factors in smart cities: a study of off-street parking in Mainz, Germany. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVI-4/W1-2021*, 97–101. https://doi.org/10.5194/isprs-archives-XLVI-4-W1-2021-97-2021

Roussel, C., Rolwes, A., & Böhm, K. (2022). Analyzing Geospatial Key Factors and Predicting Bike Activity in Hamburg. In S. Bourennane & P. Kubicek (Eds.), *Lecture Notes on Data Engineering and*

Communications Technologies. Geoinformatics and Data Analysis (Vol. 143, pp. 13–24). Springer International Publishing. https://doi.org/10.1007/978-3-031-08017-3_2

Schimohr, K., & Scheiner, J. (2021). Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption. *Journal of Transport Geography*, *92*, 103017. https://doi.org/10.1016/j.jtrangeo.2021.103017

Velleman, P. F., & Welsch, R. E. (1981). Efficient Computing of Regression Diagnostics. *The American Statistician*, *35*(4), 234. https://doi.org/10.2307/2683296

Verbeek, M. (2017). *A guide to modern econometrics* (Fifth edition). *Wiley custom*. Wiley.

Wagner, S., Götzinger, M., & Neumann, D. (2013). Optimal location of charging stations in smart cities: A points of interest based approach.

Wagner, S., Neumann, D., & Brandt, T. (2014). Smart city planning - Developing an urban charging infrastructure for electric vehicles. *Twenty Second European Conference on Information Systems, Tel Aviv*.

Wang, K., & Chen, Y.-J. (2020). Joint analysis of the impacts of built environment on bikeshare station capacity and trip attractions. *Journal of Transport Geography*, *82*, 102603. https://doi.org/10.1016/j.jtrangeo.2019.102603

Wang, X., Cheng, Z., Trépanier, M., & Sun, L. (2021). Modeling bike-sharing demand using a regression model with spatially varying coefficients. *Journal of Transport Geography*, *93*, 103059. https://doi.org/10.1016/j.jtrangeo.2021.103059

Willing, C., Klemmer, K., Brandt, T., & Neumann, D. (2017). Moving in time and space – Location intelligence for carsharing decision support. *Decision Support Systems*, *99*, 75–85. https://doi.org/10.1016/j.dss.2017.05.005

Zellner, A., Keuzenkamp, H. A., & McAleer, M. (2001). Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple. *Cambridge University Press*.

Zheng, Y., S., R., & Leckie, C. (2015). Parking availability prediction for sensor-enabled car parks in smart cities. *IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*.

Ziegel, E. R., & Myers, R. (1991). Classical and Modern Regression with Applications. *Technometrics*, *33*(2), 248. https://doi.org/10.2307/1269070