**Service Network Design of Bike Sharing Systems**

Patrick Vogel[a], Jan F. Ehmke[b], Dirk C. Mattfeld[a]

[a]Decision Support Group, Technische Unversitaet Braunschweig, Muehlenpfordtstrasse 23, 38106 Braunschweig, Germany

[b]Department Information Systems, Freie Universitaet Berlin, Garystrasse 21, 14195 Berlin, Germany

## Abstract

Bike sharing has recently enabled sustainable means of shared mobility through automated rental stations in metropolitan areas. Spatio-temporal variation of bike rentals leads to imbalances in the distribution of bikes causing full or empty stations in the course of a day. Ensuring the reliable provision of bikes and bike racks is crucial for the viability of these systems. This paper presents an integrated approach of mathematical optimization and intelligent data analysis to support service network design in bike sharing systems. Introducing the notion of service network design to bike sharing systems, we aim to show the usefulness of tactical planning for shared mobility systems.

Designing a service network requires the suitable aggregation of operational data as well as the anticipation of operational decisions. We present a mixed-integer programming formulation aiming at cost-efficient fill levels bikes at stations given a predefined service level for different scenarios of bike demand. Operational relocation decisions are anticipated by a dynamic transportation model yielding relocation services. Different scenarios of bike demand are considered as realizations of typical bike flows between stations in terms of time-dependent origin / destination matrices. We employ an intelligent data analysis approach to generate typical bike flows from individual trips recorded automatically in bike sharing systems. Intelligent data analysis produces spatio-temporal distributions of bike flows and helps determining typical trip purposes in combination with methods from the field of urban transportation planning.

The proposed methodology is exemplified based on two years of operational data from Vienna's "Citybike Wien". Computational experiments show how fill levels vary according to different scenarios of bike demand. Furthermore, spatio-temporal characteristics of relocation services are provided, which can support operators of bike sharing systems in the planning and implementation of relocation services.

**Keywords:** Service network design, Mixed-integer programming, Intelligent data analysis, Urban Transportation Planning, Shared mobility, Bike sharing

# 1 Bike Sharing Systems

Emerging metropolitan areas need efficient and sustainable mobility services in order to ensure their attractiveness, quality of life, and economic power. The more crowded a metropolitan area becomes, the more inefficient and expensive is the realization of trips with private vehicles. Municipalities have thus begun to implement innovative shared mobility systems such as car and bike sharing systems in order to accommodate the mobility needs of their citizens while ensuring sustainability and flexibility of transportation. Bike sharing systems (BSS) have become exceptionally popular. The number of implemented BSS is impressive; in Europe, about 400 BSS have been introduced in the last ten years (Büttner and Petersen 2011). Markets in America and Asia are catching up (Shaheen et al. 2010).

BSS provide an individual, but likewise public means of transportation for inner city trips (Midgley 2011). They are characterized by a high density of service facilities in heavily populated areas, e.g., with an average distance of 300 meters between bike stations (Büttner and Petersen 2011). Municipalities typically engage advertising companies for the operation of BSS (DeMaio 2009). Short bike rentals are often free of charge, and revenue is indirectly generated from a license to advertise on street furniture. Rental, return and maintenance processes are automated, enabling fast and easy access as well as one-way use and short rental times through unattended stations. Every trip is recorded for tracking and billing purposes.

While the usage of BSS is often simple, inexpensive and convenient from a user's point of view, the efficient and reliable design, management and operation of BSS are challenging. Demand for bike rentals varies strongly, following typical mobility patterns in the course of day and week caused by e.g. commuter, leisure or tourist trips. Furthermore, one-way rentals intensify imbalances in the distribution of bikes. Imbalances in the distribution of bikes affect the service level, i.e., the successful provision of bikes and free bike racks when demanded. Due to limited capacity at stations, rentals are impossible at empty stations, and returns are impossible at full stations. BSS operators aim to ensure a service level which is self-stipulated or stipulated by municipalities. For instance, a tendering for the Arlington BSS requests that "stations shall not be full of bicycles for more than 60 minutes during the hours of 8am - 6pm and 180 minutes during the hours of 6pm - 8am" (Zahory 2009).

Bike imbalances can be handled by means of strategic, tactical or operational planning. On the strategic level, decisions on the number, location and size of stations have to be made. Acquiring a high number of bike racks at stations increases the probability of successful returns. On the tactical level, bike fill levels at stations need to be determined that compensate varying bike demand in the course of day. High fill levels increase the probability of successful rentals, for example, while decreasing the probability of successful returns at particular

stations. On the operational level, relocation of bikes from rather full to rather empty stations helps maintaining the service level. Manual relocation with the help of a service fleet results in significant costs affecting the viability of BSS (DeMaio 2009), though. Planning levels are interdependent: reasonable sizing of stations and fill levels of bikes may reduce relocation efforts, whereas high relocation efforts may compensate insufficient sizing and fill levels. Hence, distinct optimization of the planning levels may lead to suboptimal decisions.
In this paper, we propose an integrated approach of intelligent data analysis and mathematical optimization supporting tactical service network design (SND) in BSS. The presented mathematical mixed-integer program (MIP) determines optimal target fill levels at stations by minimizing the expected costs of relocation. The MIP ensures a given service level for different scenarios of bike demand for a mid-term planning horizon. Scenarios are defined through bike flows that are modeled by time-dependent origin / destination (OD) matrices. The required information is derived from the aggregation of recorded customer trips in combination with well-known urban transportation planning approaches. We present an information model that abstracts from observed trip data by means of intelligent data analysis. The information model provides trip purposes, which allow for the generation of different scenarios of bike demand. These scenarios serve as input for SND.
The remainder is organized as follows. A literature overview on the analysis of shared mobility systems and related optimization approaches is presented in Section 2. We discuss our integrated approach comprising the information model representing trip purposes, its integration into an urban transportation planning approach and generation of demand scenarios in Section 3. This section also presents an optimization model for SND. The proposed methodology is exemplified with the help of a case study including two years of trip data from Vienna's BSS "Citybike Wien" (Section 4). Future work is the subject of Section 5.

## 2 Design, Management and Operation of Shared Mobility Systems

Design, management and operation of shared mobility systems can be supported by data analysis and optimization approaches for strategic, tactical and operational planning tasks. We propose the classification of planning tasks and corresponding data flows as shown in Fig. 1. The classification provides background information on the planning tasks and helps to clarify our perspective on the tactical planning level. Decisions on a specific planning level may have a significant impact on the decisions of the subordinate level. Note that this classification is also applicable to other shared mobility systems.

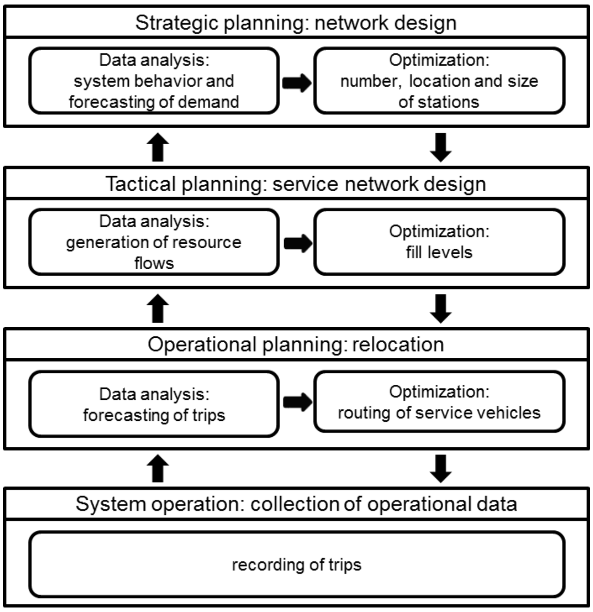


**Fig. 1** Classification of planning tasks for shared mobility systems

Following the flow of operational data, the planning levels are described from bottom to top:

- Within system operation, data about rental and return times at stations is continuously recorded for tracking and billing purposes. An exemplary trip data record comprises the particular rental station, the time of rental, the return station and the time of return.
- On the operational level, system operators need to handle short-term variation of demand by relocating bikes. Maintaining the defined service level for a specific setup of the system requires detailed routing of the service fleet. Typical decisions regard where and when to pick up and return how many bikes using which service vehicle. Due to the short-term planning horizon, routing of service vehicles relies on detailed information about current and expected bike demand, fill levels at stations, available service vehicles and staff. Based on historical trip data, short-term forecasts of trips may serve as input for the optimization of relocation operations. Forecasts have to incorporate short-term influences such as weather, events and traffic conditions.
- Tactical planning intertwines strategic and operational planning by shaping the service network. Tactical planning aims at rational and efficient management of a shared mobility system through allocation of system resources among a service network to improve the system's performance over medium-term horizons. From an optimization perspective, SND targets a certain service level in order to prepare efficient and reliable operation and relocation. Thus, decisions on SND need to take a wide-range of demand variation into account. For BSS, fill levels at stations are to be determined as input for operational planning. To optimize the SND for a given service level, tactical planning requires the suitable aggregation of operational data and the anticipation of operational decisions. Our tactical planning approach is detailed in Section 3.

- On the strategic level, typical decisions regard the number, location and (re-)size of stations ensuring sufficient overall coverage and capacity of the implemented system. The analysis of historical trip data as well as external data such as demographic and land-use data may lead to insights on user behavior. Strategic planning also allows for the determination of overall demand for network design purposes.

In the following, a literature overview of the above planning tasks is provided, especially with regard to different planning and optimization tasks for BSS. We also discuss the role of operational data and its required level of aggregation for the particular planning tasks. First, recent work on the well-studied and established operational and strategic planning levels are presented. Second, studies on the rather unacquainted and unappreciated tactical planning level are described.

### 2.1 Operational and Strategic Planning

For *operational planning*, a system operator needs to forecast trips on a very detailed, short-term level, supporting planning of relocation tours. Statistical analysis of observed trip data may provide the required data input. Vogel and Mattfeld (2011) model and forecast bike rentals on an hourly basis while considering seasonal influences by detailed weather data. Borgnat et al. (2010) incorporate weather data and event data (holidays, strikes) for short-term forecasts of bike rentals. Kaltenbrunner et al. (2010) and Froehlich et al. (2009) analyze fill levels in order to forecast the availability of bikes at stations.

Many authors relate the optimization of relocation tours to the one-commodity-pickup-and-delivery problem (PDP) and the swapping problem (SP). In the PDP, a fleet of vehicles transports a commodity from pickup to delivery stations. In the SP, multiple commodities are considered and a station serves both as a pickup and as a delivery station. Planning of relocation services is studied as a static or dynamic problem. For the static problem, relocations are realized at night time when no demand occurs. Benchimol et al. (2011) combine PDP and SP and present a static model and solution methods. Raviv et al. (2013) study the static relocation problem minimizing user dissatisfaction by means of penalty costs and operating costs for relocation. Ricker et al. (2012) introduce a simulation-based approach to determine the cost-efficient daily number of relocation operations. Weighted sums of transportation costs and costs for unserved customers are considered. Rainer-Harbach et al. (2013) propose a variable neighborhood search in combination with a greedy heuristic, maximum flow approach and linear program (LP) to determine the routes and number of relocated bikes for the static relocation problem. Raidl et al. (2013) improve the variable neighborhood search of the former work by efficiently determining optimal loading operations. Also addressing the static case, Di Gaspero et al. (2013a) present a hybrid metaheuristic combining constraint-based programming and ant colony optimization. The objective is to minimize the travel time for relocation tours and the difference between actual and target fill levels at stations. Di Gaspero et al. (2013b) extend the constraint-based programming approach and incorporate large neighborhood search to speed up to the branching strategy inherent in the constraint programming. Ho and Szeto (2014) study the static relocation problem and propose an iterated tabu search.

For the dynamic problem, demand variation and several decision points over time are considered. Contardo et al. (2012) present an arc-flow optimization model for the dynamic routing of service vehicles minimizing “lost demand”. Lost demand is caused by customers who cannot rent or return bikes at empty or full stations, respectively. Caggiani and Ottomanelli (2012) propose a decision support system for the dynamic relocation problem. Here, a neural network is used to forecast rentals and returns at stations. Dell’Amico et al. (2013) develop MIP formulations for the dynamic relocation problem based on the one-commodity pickup and delivery capacitated vehicle routing problem. Since the formulations lead to an exponential number of constraints, a branch-and-cut algorithm is introduced. Kloimüllner et al. (2014) extend the previous work of Rainer-Harbach et al. (2013) and Raidl et al. (2013) to the dynamic case. They use a greedy construction heuristics and present two metaheuristic approaches, namely greedy randomized adaptive search and variable neighborhood search. Kaspi et al. (2014) simulate different parking reservation policies in shared mobility systems based on a Markov chain model. Upon rental, users specify their destination and parking spaces are reserved according to different policies. The simulation shows that reservation policies reduce the travel time of users. Fricker and Gast (2014) apply a Markov model to simulate user trips in BSS. They show that simple incentives, e.g. suggesting users to return the bike at the station with the lowest fill level among two stations, can improve the performance of the BSS.

For car sharing systems, Kek et al. (2009) present a MIP minimizing costs for service staff and relocation operations. Lost demand is considered in terms of penalty costs. Nair and Miller-Hooks (2011) propose a stochastic MIP with chance constraints to obtain a least-cost plan for the relocation of vehicles. Their cost function comprises fixed costs for the relocation of vehicles, relocation between stations, and penalty costs for the utilization of additional service vehicles. Di Febbraro et al. (2012) apply simulation for user-based relocation. Discounts are given if users return rental cars at locations proposed by an assignment model matching cars and rental demand. Weikl and Bogenberger (2012) present a conceptual framework for relocation in free-floating car sharing systems. In their two-step approach, they first identify clusters having similar demand patterns and second apply relocation strategies based on the clusters. Nourinejad and Roorda (2014) maximize the total profit of a one-way car sharing systems considering revenue for trips and costs for relocation. Their proposed decision support system comprises discrete event simulation and the optimization of relocations.

For *strategic planning*, information on typical system behavior is required. Based on data analysis of a large data set of customer trips, Vogel et al. (2011) determine temporal demand "activity clusters" describing typical rental and return activities at stations in the course of the day. Cluster analysis reveals groups of stations with similar trip purposes represented by the activity. Borgnat et al. (2010) characterize interrelated stations by cluster analysis of bike flows between stations. Wang et al. (2012) apply linear regression to model the correlation of bike activity at stations and external factors like demography and transportation infrastructure. Faghih-Imani et al. (2014) model the influence of weather, bike infrastructure, land-use and environmental attributes on bike rental and return rates. O'Brien et al. (2013) present various key figures on the configuration of BSS as well as their spatial and temporal characteristics.

Consideration of spatial relations between bike rentals at stations and location of stations may support strategic decisions on the number, location and size of stations. Lin and Yang (2011) present a hub-location model that determines the number and locations of bike stations as well as the network of bike paths. Here, customers' travel costs and setup costs for bike stations and bike paths are minimized. In an extended version of their optimization model, also decisions on the bike inventory at stations are taken into account (Lin et al. 2013). Martinez et al. (2012) propose a MIP to optimize the location of bike sharing stations and the size of the bike fleet. Garcia-Palomares et al. (2012) introduce a location-allocation modeling approach to optimize the location of bike sharing stations based on coverage. Nair and Miller-Hooks (2014) present a MIP for the optimal configuration of shared mobility systems by determining the station locations and sizes as well as vehicle inventories, but neglect operational decisions. Chow and Sayarshad (2014) propose an approach for the integrated network design of BSS and traditional public transportation systems.

### 2.2 Tactical Planning

Compared to work on strategic and operational planning of shared mobility systems, literature on tactical planning is scarce. Existing studies handle tactical planning with and without anticipation of operational decisions. The following studies do not anticipate operational decisions:

- George and Xia (2011) model shared mobility systems by means of a closed queuing network. A profit maximizing optimization is applied in order to determine the optimal fleet size and allocation of rental vehicles.
- Cepolina and Farina (2012) determine the fleet size and vehicle allocation for a car sharing system with small electric vehicles. Costs for user waiting times and system operation (vehicle purchasing and running costs) are minimized by means of Simulated

Annealing. Dynamic user-based relocation is assumed to be coming at no additional cost. Thus, relocation costs are not taken into account.

- Raviv and Kolka (2013) also use queuing models. With the help of a user dissatisfaction function, the optimal fill level at a bike station is determined.
- Schuijbroek et al. (2013) minimize the costs of relocation tours and incorporate service level requirements at stations. They consider the static case in which no varying user demand is considered. The service level is precalculated for each station without anticipation of the routing decisions. A cluster-first route-second heuristic is proposed to solve the problem.
- Shu et al. (2013) use a network flow model to determine the initial allocation of bikes at stations in order to maximize bike flows and successful trips within the network on weekly basis. In a separate optimization model, they assess the impact of relocations on the number of required bikes in the system.

Especially for tasks of tactical planning, anticipation of operational decisions is crucial for the viability of shared mobility systems. Costly relocation could be alleviated by appropriate fill levels that compensate expected variation of demand. To the best of the authors' knowledge, only two integrated studies exist, anticipating relocation operations in tactical planning:

- Correia and Antunes (2012) present multi-periodic MIP formulations to maximize the profit of a car sharing system considering the revenue of trips, costs of depot and vehicle maintenance as well as costs of vehicle relocation. They determine the number and the location of stations as well as the number of vehicles at each station in each period of daily operation. They consider static relocation at the end of the day where vehicles are relocated between stations to reset the initial fill level. The validity of the MIP approach is investigated by means of a simulation model (Jorge et al. 2012).
- Sayarshad et al. (2012) introduce a dynamic LP formulation to maximize profit in BSS. Relocation, maintenance, capital and holding costs of bikes as well as penalty costs for lost demand are deducted from revenue generated by trips. Unutilized bikes can be relocated in every period of daily operation.
- Boyaci et al. (2015) present an optimization framework for the development of CSS. In a MIP formulation, the revenue of the CSS is maximized taking strategic, tactical and operational decisions into account. Due to the high number of relocation variables, an imaginary hub station is introduced. Relocation is only considered between bike stations and the hub station significantly reducing the number of relocation variables.

In sum, recent approaches of tactical planning do not sufficiently reflect the interaction of fill levels and relocation operations as known from the field of SND. A general methodology that benefits from usage and aggregation of detailed operational trip data for tactical planning is missing. Thus, in the following, we adapt existing optimization approaches of SND focusing on the adequate anticipation of relocation tours and present a new approach to aggregate operational data as input for SND.

# 3 Service Network Design for Bike Sharing Systems

SND requires the aggregation of operational data and the anticipation of operational decisions. In this section, for data aggregation, an information model is proposed, which represents typical bike flows for different scenarios of bike demand by time-dependent OD matrices (cf. Section 3.1). In combination with approaches from the field of urban transportation planning, trip purposes are identified, which allow for the generation of bike flows as input for SND. In Section 3.2, an optimization model is presented based on a MIP formulation, aiming at cost-efficient allocation of bikes to stations while ensuring a predefined service level for different scenarios of bike demand. We determine the total number of bikes in the system, optimal target fill levels of stations, and expected relocation operations. Target fill levels ensure the provision of service depending on the time of the day for a given scenario, e.g., high bike demand on a working day in the main season. The anticipation of relocation operations yields the expected costs of relocation services to compensate insufficient fill levels.

## 3.1 An Information Model for Generation of Typical Bike Flows

BSS automatically record extensive amounts of trip data. Recorded trip data represent individual observations of customer behavior and are therefore not suited as input for tactical planning. Thus, we propose a combined approach of urban transportation planning and intelligent data analysis to derive an information model that represents trip purposes and typical bike flows (cf. Section 3.1.1). Based on this concept, we detail how this information model can be used to generate bike flows for SND (cf. Sections 3.1.2).

### 3.1.1 Combining Transportation Planning and Intelligent Data Analysis

To fully explore spatio-temporal characteristics of bike trips, we align the Urban Transportation Planning Systems (UTPS) process (Johnston 2004) with the field of intelligent data analysis (Berthold et al. 2010). The UTPS process is a common approach to model trips in urban areas. It comprises trip generation, trip distribution, mode choice and route selection, and provides an estimate of traffic flows for individual links of the considered transportation network. Intelligent data analysis refers to the non-automatable extraction of knowledge from large datasets by means of data driven methods such as data mining (Berthold et al. 2010). Input data for the UTPS process is usually derived from costly surveys. Based on a small sample size, surveys provide a general picture of traveler behavior, including background information on the purpose of a trip and the traveler's attitude towards the usage of shared mobility systems. Extending this idea with approaches from intelligent data analysis, we derive trip purposes from an extensive amount of trip data recorded by BSS. Trip data are available at low costs, since they have already been collected for tracking and billing purposes. The challenge is to derive the different trip purposes from the observed rental and return operations by intelligent data analysis. Once this is done, typical measures of mobility behavior and different demand scenarios can be generated for different demand scenarios.

Although trip data are a great source for tactical planning, there are also limitations. First, observed trip data may be biased, since lost demand is not recorded. As discussed in Section 2, lost demand describes situations where a user would have rented a bike if the station had not been empty. However, this is not contained in recorded trip data. User polls or surveys asking for the general mobility behavior may give some indication of "real" demand, but again, these approaches usually rely on small samples and may also be biased (Flyvbjerg et al. 2006). Second, relocation operations already carried out by the service operator affect the characteristics and the number of realized trips. Filtering the effects of past relocation operations is nearly impossible due to complex spatio-temporal interdependencies, though. For instance, simply erasing trips that might have not been realized is too short-sighted, since they might have affected the fill levels of a chain of stations in the course of a day. Thus, for the scope of SND as described in this paper, we assume that tactical planning does not change customer behavior immediately, and that past relocation operations did not counteract the mobility demand of users.

#### 3.1.2 Generation of Typical Bike Flows

Aggregation of trip data is required to model the spatio-temporal distribution of trips and underlying trip purposes. We employ intelligent data analysis to segment bike stations according to varying rental and return activities in the course of the day. With the temporal distribution at hand, the spatial distribution of trips between groups of stations with similar temporal activity are then determined.

***Construction of the information model.*** The temporal segmentation of bike stations aims to provide a compact representation of demand variation for SND. The idea is to represent typical demand without smoothing out information about demand variation. To this end, *temporal activity clusters* are constructed by cluster analysis. A temporal activity cluster refers to a group of stations with similar rental and return activities in the course of the day. According to preliminary analyses of trip data, hourly aggregation seems suitable for cluster analysis. We implement the Expectation-Maximization algorithm (Dempster et al. 1977) to provide temporal activity clusters as detailed in Vogel et al. (2011). Results of cluster analysis are evaluated by several internal validation measures as well as external validation by exploratory data analysis and interviews with operators of the BSS. As a result, each station is characterized by its assigned temporal activity cluster, which yields the typical proportion of rentals and returns for each hour of the day.

The spatial distribution of trips between stations is derived from the associated temporal activity clusters as follows:

- First, the *inter-cluster distribution* is constructed, which describes trip distribution patterns between stations of individual activity clusters. They are specified by the proportion of trips between individual temporal activity clusters for a given hour of the day. For instance, in the morning, the majority of trips is directed from “residential” to “working” clusters, whereas the opposite is true for afternoon hours.
- Second, the *intra-cluster distribution* specifies how trips are distributed from a particular station to all stations contained in a cluster. We approximate this distribution based on the distance between stations and the resulting trip duration. The distribution of trip durations is empirically derived from recorded trip data.

***Formalization of the information model.*** With the temporal and spatial distributions at hand, the information model can be formalized as follows:

- The BSS consists of a set bike stations $N = \{s_1, \dots, s_n\}$.
- The planning horizon comprises $T = \{0, .., t_{max}\}$ periods, e.g., 24 hourly periods representing a typical working day.
- The total activity of a station $s_i$ is denoted by the absolute number of daily rentals $B_i^-$ and daily returns $B_i^+$.
- The set of temporal activity clusters is $C = \{c_1, \dots, c_z\}$.

The clustering $\gamma: N \to C$ assigns each station $s_i \in N$ to a temporal activity cluster $c_j \in C$ defining the trip purposes at the station. Trip purposes are represented by the temporal rental activity $\beta_{c_j,t}^- \in [0,1]\ \forall t \in T, c_j \in C$. The temporal activity expresses the relative hourly activity and thus summarizes to 1 over the course of the day for each cluster, i.e., $\sum_{t \in T} \beta_{c_j,t}^- = 1\ \forall c_j \in C$. The same holds for returns $\beta_{c_j,t}^+ \in [0,1] \forall t \in T, c_j \in C$ with $\sum_{t \in T} \beta_{c_j,t}^+ = 1\ \forall c_j \in C$.

The spatial trip distribution is given according to the inter-cluster distribution $\kappa: C \times C \times T \to [0,1]$ and intra-cluster distribution $\lambda: N \times N \to [0,1]$:

- The inter-cluster distribution expresses the fraction of flows between clusters per time period. The fraction of inter-cluster flows summarizes to 1 from a particular cluster $c_i$ in a specific time period $t$ to all clusters $c_j$ by means of $\sum_{c_j \in C} \kappa_{c_i c_j,t} = 1\ \forall c_i \in C, t \in T$.
- The intra-cluster distribution expresses the fraction of flows from station $s_i$ to station $s_k$ depending on the assigned cluster. The fraction of intra-cluster flows summarizes

to 1 based on flows $s_i$ to all stations $s_k$ of the particular cluster by means of $\sum_{s_k \in C_j} \lambda_{s_i s_k} = 1 \ \forall C_j \in C, s_i \in N$.

With these notations in mind, we can describe the temporal and spatial distribution of bike rentals as follows:

1) Temporal distribution: We determine the hourly activity at stations $B_{i,t}^{-}$ by distributing the number of rentals at stations to the time periods provided by the temporal rental activity:

$$B_{s_i,t}^{-} = B_{s_i}^{-} \cdot \beta_{\gamma(s_i),t}^{-} \ \forall s_i \in N, t \in T$$

2) Spatial distribution:
   a) Inter-cluster distribution: We determine the bike flows $f_{s_i c_j,t}^{-}: N \times C \times T \rightarrow \mathbb{R}^{+}$ from each station to each cluster by distributing the hourly rentals to the clusters:

   $$f_{s_i c_j,t}^{-} = B_{s_i,t}^{-} \cdot \kappa_{\gamma(s_i) c_j,t} \ \forall s_i \in N, c_j \in C, t \in T$$

   b) Intra-cluster distribution: We determine the bike flows $f_{s_i s_j,t}^{-}: N \times N \times T \rightarrow \mathbb{R}^{+}$ from each station $s_i$ to each station $s_j$ by distributing the bike flows to the clusters among the stations belonging to the clusters:

   $$f_{s_i s_j,t}^{-} = f_{s_i \gamma(s_j),t}^{-} \cdot \lambda_{s_i s_j} \ \forall s_i, s_j \in N, t \in T$$

The distribution of bike returns can be determined analogously. In the end, rental and return flows are averaged. Output of the information model are time-dependent, real-valued bike flows $f_{s_i s_j,t}$, which represent the expected bike flow between origin station $s_i$ and destination station $s_j$ in hour $t$.

***Example generation of bike flows.*** Let us clarify the generation of bike flows based on a numerical example. In a specific hour of the day ($t = 8$), we consider rentals at station $s_i$ in cluster $\gamma(s_i) = c_x$ with a particular rental activity $\beta_{c_x,t}^{-}$ and two destination clusters $c_y$ and $c_z$ (cf. Fig. 2). Cluster $c_y$ contains the stations $s_j$ and $s_k$. Station $s_j$ is closer to station $s_i$ than $s_k$. For the sake of simplicity, the stations of $c_z$ are not considered. The expected number of daily rentals for station $s_i$ is $B_{s_i}^{-} = 100$.

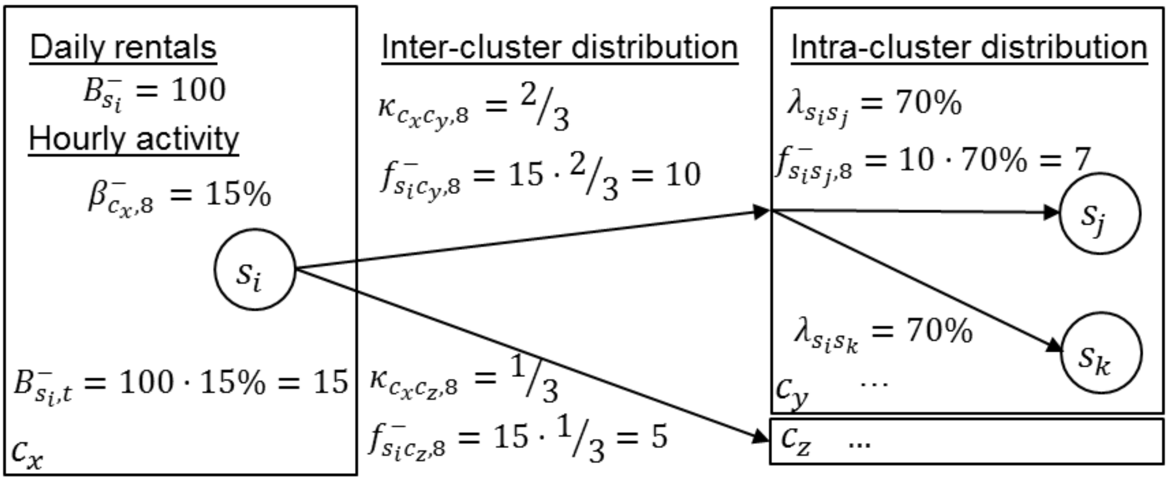


**Fig. 2** Flow generation according to the temporal activity and spatial distribution

Combining the temporal and spatial distributions, the flows from station $s_i$ to $s_j$ and $s_k$ are computed as follows:

1) Temporal distribution: The number of daily rentals is temporally distributed according to the rental activity cluster of this station, which denotes that 15% of the daily rentals account for the considered hour of the day.
2) Spatial distribution:
   a) Inter-cluster distribution: Rentals are spatially distributed to activity clusters according to the inter-cluster distribution for the given hour of the day. Here, 2/3 of rentals are distributed from cluster $c_x$ to cluster $c_y$ and 1/3 from cluster $c_x$ to cluster $c_z$. Hence, 10 rentals are distributed to $c_y$ and 5 to $c_z$.
   b) Intra-cluster distribution: Rentals are further distributed within each activity cluster according to the intra-cluster distribution, contributing to a particular flow between a given OD pair. For this example, we assume that 70% of the trips have a short duration and 30% have a long duration. Thus, 70% of the rentals from station $s_i$ are assigned to station $s_j$ and 30% to station $s_k$, because $s_j$ is closer to $s_k$.

In sum, for the given hour of the day, the generated flow from station $s_i$ to $s_j$ is 7 bikes and from station $s_i$ to $s_k$ is 3 bikes. Since each station also serves as an attractor of trips and the number of daily rentals and returns may differ, the procedure is executed again to compute the number of bike rentals at each station. Let us assume that the generated flow to station $s_i$ from $s_j$ is 4 bikes and to station $s_i$ from $s_k$ is 1 bike. In the end, the average of the generator and attractor bike flows is computed, providing time-dependent OD matrices. Here, the average flow for the 8$^{th}$ hour of the day from station $s_i$ to $s_j$ is 5.5 and from station $s_i$ to $s_k$ is 2 bikes.

***Generation of integer flows.*** The above presented method provides typical bike flows for each pair of stations in each time period in terms of *real-valued* metrics. However, realistic anticipation of relocation operations requires *integer-valued* bike flows, since it does not make sense to anticipate fractions of bike relocations. To this end, we transform the real-valued bike flows into integer bike flows by scaling and transformation as follows:

- In the scaling step, the real valued bike flows $f_{s_i s_j,t}$ are multiplied such that they equal the desired number of bike flows $d$ in relation to the total number of observed bike flows $o$ with $f'_{s_i s_j,t} = mult \cdot f_{s_i s_j,t} \forall s_i, s_j \in N, t \in T$, $mult = {}^{d}/_{o}$.
- In the transformation step, the flows are rounded according to a threshold $\tau$ for rounding up and down such that the total number of rounded flows amounts to the desired number of flows: $\sum_{s_i \in N} \sum_{s_j \in N} \sum_{t \in T} Round\left(\tau , f'_{s_i s_j,t}\right) = d$ with $Round\left(\tau, f'_{s_i s_j,t}\right): \ If\ f'_{s_i s_j,t} - \left\lfloor f'_{s_i s_j,t} \right\rfloor < \tau\ then \left\lfloor f'_{s_i s_j,t} \right\rfloor\ else\ \lceil f'_{s_i s_j,t} \rceil$.

A binary search is applied to determine $\tau$ yielding the desired number of bike flows. As a result, we have provided integer-valued, time-dependent OD matrices of bike flows as required for SND as detailed below.

### 3.2 MIP Formulation for SND

The following optimization model is based on the work of Crainic (2000) on SND in freight transportation. Generally, decisions on the tactical level aim at the optimal allocation and utilization of resources to fulfill customer service and economic goals. Total costs comprise fixed costs for offering a regular transportation service between particular locations in a network and variable costs that arise for a particular set of transported goods. Transferring the idea of SND to the area of BSS, the service operator transports bikes in capacitated trucks from full to empty stations as to maintain a given service level. Fixed transportation costs arise for the implementation of relocation services, and variable costs arise for the handling of transported bikes.

We propose a MIP formulation which determines optimal target fill levels at stations in the course of the day, ensuring the fulfillment of demand scenarios according to a predefined service level. The objective is to obtain fill levels at minimal expected costs of system operation. Resulting target fill levels and relocation services may serve as input for the optimization of relocation tours on the operational level. Relocation tours have to be adjusted depending on the actual demand realization of the particular day.

Within the scope of tactical planning, anticipation of operational decisions is required to avoid suboptimal decisions on fill levels. Our optimization model is based on a relaxation of relocation operations. We refrain from a detailed modeling of routing as known from traditional computationally challenging SND models (Crainic 2000) or inventory routing models (Campbell et al. 1998), but we anticipate relocation operations by means of a dynamic transportation model (Bookbinder and Sethi 1980) yielding the required demand for relocation services. To this end, we use a binary variable allowing constraints on the frequency and the capacity of relocation services by consolidating relocations.

A relocation service is described by pickup and return station, time period, and the number of relocated bikes. Relocation services represent the design decision for implementing a service between two stations in each period at each day of system operation. They are modeled by binary variables $RS_{s_i s_j,t}$. The number of relocated bikes for a particular service is modeled by continuous variables $R_{s_i s_j,t}$. Again, note that this is an approximation of relocation operations, which cannot replace detailed optimization from an operational perspective by means of vehicle routing procedures.

Let $N$ be a set of rental stations and $T$ the set of time periods in a day. The total number of bikes in the system is given by $b$. The number of bikes can be adjusted if needed, e.g., in case of transition into another season. The typical demand for bikes and bike racks is depicted by bike flows $f_{s_i s_j,t}$ between stations $s_i$ and $s_j$ in time period $t$. The fulfillment of demand at stations depends on the given design and configuration of system infrastructure, i.e., the number of bike racks $br_{s_i}$ for returns ("size" of a station) and the number of allocated bikes at each station and period $B_{s_i,t}$ for rentals. The objective is to minimize the total costs for relocation services while ensuring the availability of rental and return resources for time-dependent "safety buffers" of bikes $sb_{s_i,t}$ and bike racks $sbr_{s_i,t}$. Based on time-dependent OD matrices, the information model provides a scenario of bike flows $f_{s_i s_j,t}$ that serve as input for optimization.

The SND model reads as follows:

Sets

- $N = \{s_1, \dots, s_n\}$ : set of bike stations
- $T = \{0, .., t_{max}\}$ : set of time periods, e.g., hours of the day. For resetting the number of allocated bikes at the end of the day, $t_{max}$ includes the first period of the next day.

Decision variables

- $B_{s_i,t} \in \mathbb{R}$ : number of bikes at station $s_i$ in time period $t$
- $R_{s_i s_j,t} \in \mathbb{R}$ : number of relocated bikes between stations $s_i$ and $s_j$ in time period $t$
- $RS_{s_i s_j,t} \in \{0,1\}$ : defines whether there is a relocation service between stations $s_i$ and $s_j$ in time period $t$

Parameters

- $br_{s_i}$ : number of bike racks at station $s_i$ (size of a station)

- $b$ : total number of bikes in the system
- $f_{s_i s_j,t}$ : bike flow between stations $s_i$ and $s_j$ in time period $t$
- $ch_t$ : average handling costs of one relocated bike in time period $t$
- $ct_{s_i s_j}$ : average transportation costs of one relocated bike between stations $s_i$ and $s_j$
- $l$ : lot size, defining the capacity of the relocation truck
- $sb_{s_i,t}$ : bike safety buffer at station $s_i$ in time period $t$
- $sbr_{s_i,t}$ : bike rack safety buffer at station $s_i$ in time period $t$

With this notation, the optimization model reads:

$$\text{Minimize} \quad \sum_{t=0}^{t_{max}} \sum_{s_i=1}^{n} \sum_{s_j=1}^{n} \left( ch_t \cdot R_{s_i s_j,t} + ct_{s_i s_j} \cdot RS_{s_i s_j,t} \right) \quad (1)$$

subject to

$$l \cdot RS_{s_i s_j,t} \geq R_{s_i s_j,t} \; \forall \mathrm{s}_i, \mathrm{s}_j \in N, t \in T \; (2)$$

$$B_{s_i,t+1} = B_{s_i,t} + \sum_{s_j=1}^{n} (f_{s_j s_i,t} - f_{s_i s_j,t} + R_{s_j s_i,t} - R_{s_i s_j,t}) \; \forall \mathrm{s}_i \in N, t \in T \backslash t_{max} \quad (3)$$

$$B_{s_i,t} - \sum_{s_j=1}^{n} f_{s_i s_j,t} + \sum_{s_j=1}^{n} f_{s_j s_i,t} - \sum_{s_j=1}^{n} R_{s_i s_j,t} \geq \mathrm{sb}_{\mathrm{s_i,t}} \; \forall \mathrm{s}_i \in N, t \in T \quad (4)$$

$$\mathrm{br}_{\mathrm{s_i}} - B_{s_i,t} - \sum_{s_j=1}^{n} f_{s_j s_i,t} + \sum_{s_j=1}^{n} f_{s_i s_j,t} - \sum_{s_j=1}^{n} R_{s_j s_i,t} \geq \mathrm{sbr}_{\mathrm{s_i,t}} \; \forall \mathrm{s}_i \in N, t \in T \quad (5)$$

$$R_{s_i s_j,0} = 0 \; \forall \mathrm{s}_i, \mathrm{s_j} \in N \quad (6)$$

$$B_{s_i,0} = B_{s_i,t_{max}} \; \forall \mathrm{s}_i \in N \quad (7)$$

$$\sum_{s_i=1}^{n} B_{s_i,t} = \mathrm{b} \; \forall \mathrm{t} \in \mathrm{T} \quad (8)$$

$$B_{s_i,t}, R_{s_i s_j,t} \geq 0 \; \forall \mathrm{s}_i, \mathrm{s}_j \in N, t \in T \quad (9)$$

In the objective function (1), the costs for anticipated relocation services are minimized, comprising handling costs for each individual bike $R_{s_i s_j,t}$ and setup costs for running the particular relocation service $RS_{s_i s_j,t}$ between two stations. Handling costs can vary depending on the time of the day, e.g., there are higher costs at night due to surcharges for the staff. Transportation costs are assumed to be constant. Depending on the given infrastructure configuration, potentially missing bikes or bike racks are compensated by relocation of bikes $R_{s_i s_j,t}$ between stations for each period of the day. Constraint (2) ensures that a relocation service does not exceed a predefined capacity given by the lot size $l$. Equation (3) ensures flow conservation, i.e., the number of bikes at a station in the next period is a result of the current number of bikes plus returns from customers ($f$) and relocation services ($R$) minus customer rentals and relocation pickups. We assume that a particular relocation service is realized within one time period, but if relocation services take longer, (2) has to be adjusted by setting $R_{s_j s_i, t-1}$ as well as the range of the index $t$.

The availability of resources is ensured by constraints (4) and (5). On the one hand, it is guaranteed that a sufficient number of bikes (4) is present at every station and period, i.e., the number of bikes minus customer rentals plus costumer returns and relocation pickups is always larger than a given bike safety buffer $sb_{s_i,t}$. On the other hand, the number of free bike racks (bikes racks minus allocated bikes, customer and relocation returns plus customer rentals) is always larger than the bike rack safety buffer $sbr_{s_i,t}$ (5). These two constraints ensure that rented bikes and used bike racks are not available for relocation in the particular period, and all demand is satisfied. Relocation services are not allowed in the first period (6), and the initial fill level is restored at the end of the day (7). Equation (8) ensures that all existing bikes need to be allocated. Decision variables must be non-negative (9). The above constraints enable particular safety buffers for bike and bike racks depending on the time of day. For instance, in periods with a high rental activity and a low return activity at a station, the bike safety buffer can be set to a high value while the safety buffer can be kept low for bike racks. Reasonable values for safety buffers can be determined by analyzing the demand variation based on observed trip data.

Modeling the availability of resources as shown in constraints (4) and (5) is a rather optimistic approach, since customer rentals and returns are interchanged simultaneously. An alternative approach would be to handle bikes and bike racks as separate resources, but this would result in a too pessimistic modeling since recently returned bikes could not be used by the next customer in the same time period.

Although we relax the construction of relocation tours by applying a dynamic transportation model, the introduced dynamic MIP formulation is still computationally hard. The complexity arises due to the large number of binary variables for relocation services (stations × stations × time periods). We note that the MIP model will be computationally intractable for large instances, and that heuristics may be required here.

## 4 Decision Support for Service Network Design of "Citybike Wien"

In the following study, the presented approach is applied to a real BSS in order to demonstrate the usefulness of SND and the interplay of information and optimization models. The information model is parameterized based on extensive trip data recorded by Vienna's "Citybike Wien". Two demand scenarios are generated (Section 4.1). For each scenario, results of SND are discussed along spatio-temporal dimensions (Section 4.2).

### 4.1 Generation of Typical Bike Flows for Service Network Design

Citybike Wien provided trip data for the years 2008 and 2009. The operational dataset comprises approx. 750’000 data records for a BSS of 59 stations with a total of 1253 bike racks and 627 bikes. In order to employ a tactical planning perspective and to reflect the typical usage of the system, we restrict our analysis to summer trips only (April to October), accounting for 72% of all trips. In the summer season, 1569 trips occur per day or 2.5 trips per bike and day on average, respectively. The data analysis tool RAPIDMINER (http://rapid-i.com/) is used for generation, documentation and implementation of the information model. The transformation and the scaling of flows is implemented in JAVA.

#### 4.1.1 Temporal Distribution of Trips

For temporal modeling, we need to determine an appropriate timescale in order to aggregate operational trip data for SND. Based on recorded rental and return times, the durations of trips are calculated and then aggregated in one minute buckets. Trip durations follow a Poisson-like distribution (Fig. 3a). The average trip lasts approx. 27 minutes with a median of 16 minutes. About 92% of trips are shorter than 60 minutes. Almost 70% of the trips end within the same hour, i.e., a trip that starts in a particular hour of the day will most likely end in the same hour. For this set of operational data, hourly aggregation seems sufficient to reflect temporal variation of bike rentals. The distribution of trip durations combined with given distances between stations determine parameters for the intra-cluster distribution $\lambda_{s_i s_j}$.

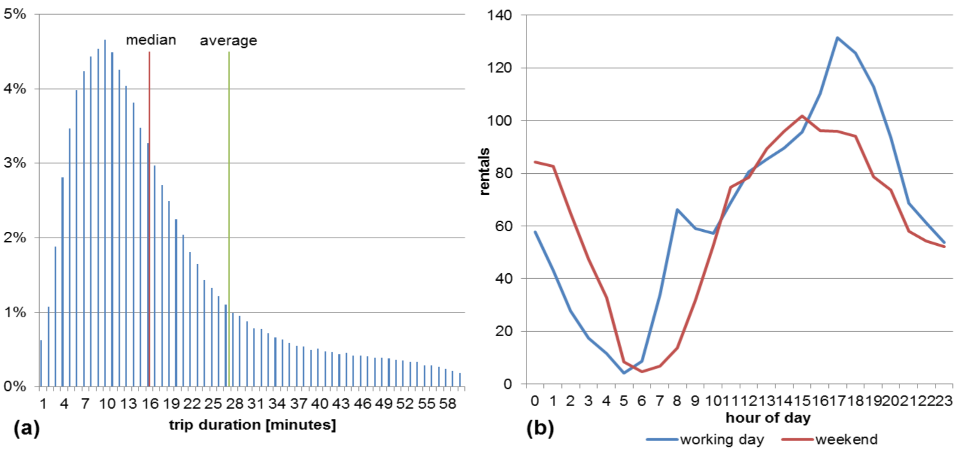


**Fig. 3** Distribution of trip durations **(a)** and rental patterns on working and weekend days **(b)**

The hourly aggregation of trips results in 24*7 values representing the average rental activity in every hour of each day. Working days and weekend days show distinct rental patterns (Fig. 3b). Working days have three peaks: (1) a night peak probably resulting from the cessation of subway service, (2) morning commutes, and (3) the overall daily peak in the afternoon hours due to overlapping commuter and leisure usage. Weekend days clearly indicate a leisure-dominated activity by a distinct night peak and missing morning peak. In the following, we focus on the analysis of working days only, since relocation is usually not carried out on weekends. To this end, we set $t_{max} = 24$ (hourly) periods.

In order to determine temporal activity cluster of stations, the hourly rental and return activity for each station is calculated, i.e. the fraction of daily rentals and returns, respectively. This leads to a data set of 59 stations with 48 attributes representing the temporal activity. For the above data set, cluster analysis groups the 59 stations to five activity clusters. Cluster centroids represent the main trip purposes at stations that were assigned to the particular cluster.

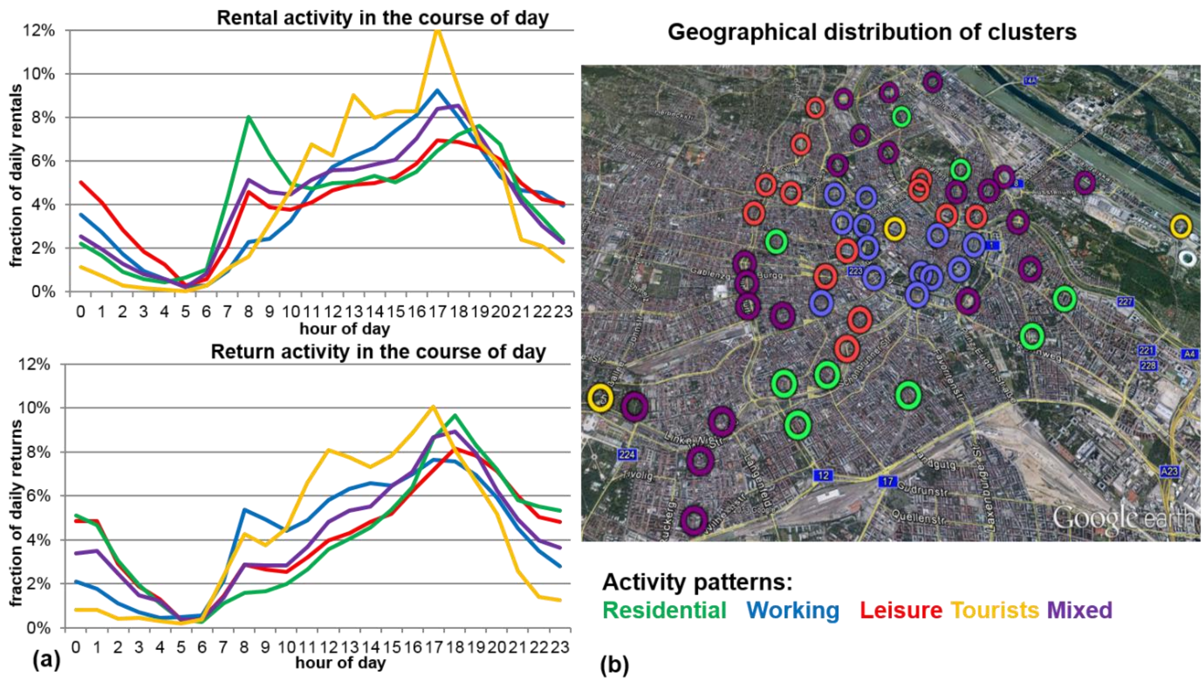


**Fig. 4** Rental and return activity clusters **(a)** and geographical distribution of clusters **(b)**

Fig. 4 shows the obtained rental activities $\beta^-_{C,t}$ and return activities $\beta^+_{C,t}$ as well as the geographical distribution of clusters in the city of Vienna:

- Stations within the *working* cluster are characterized by commuter trips showing a return activity peak in the morning and a rental activity peak in the late afternoon. These stations are located in the city center, having a high number of working places and points of interest as well as a low proportion of residents.
- The *residential* cluster shows the opposite activity of commuter trips with dominating rental activity in the morning and return activity in the afternoon. These stations are located at the periphery, which has more residential buildings.
- The *leisure* cluster shows activities similar to the residential cluster, but stands out due to different nighttime activities likely resulting from leisure trips. These activities are probably caused by popular nightlife districts.

- The *tourist* cluster is distinguished by a significant proportion of daytime rental and return activity, but almost no nighttime activity. Stations are close to popular tourist attractions in the west (castle Schoenbrunn), east (Prater carnival) and the city center (St. Stephan's Cathedral). Note that Citybike Wien's "tourist card" is also handed out next to the city center station, which may explain the distinguished activity of this station.
- The *mixed* cluster represents stations that cannot be distinguished according to their main trip purposes and thus reflects a more average rental and return activity on working days. This observation is also underlined by the location of these stations, which is often between stations of other clusters.

#### 4.1.2 Spatial Distribution of Trips

Based on the temporal activity clusters, the spatial distribution of trips between temporal activity clusters is computed considering the different trip purposes (working, residential, leisure, tourist, mixed). We exemplify the results for the time-dependent inter-cluster distribution $\kappa_{CC,t}$ for stations of the residential cluster (cf. Fig. 5).

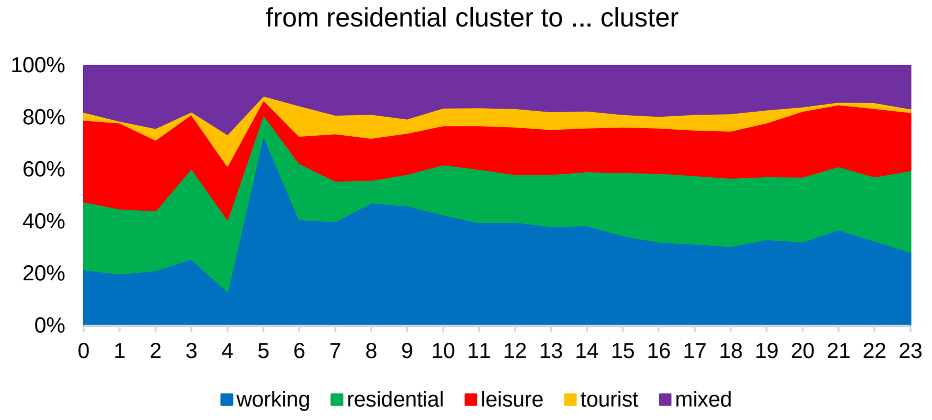


**Fig. 5** Inter-cluster distribution between the residential cluster and other clusters

In the morning hours, more than 40% of trips starting at the residential cluster end at the working cluster reflecting commuter trips. Note that the peak in hour 5 with of proportion of 70% commuter trips might be overrepresented, since this is the hour with the lowest overall usage. In the afternoon hours, the proportion of trips from the residential cluster to working cluster declines. In contrast, the proportion of trips to the residential and leisure cluster increases. Trips to the leisure cluster dominate during night time. In sum, the inter-cluster distribution follows the general mobility behavior in Citybike Wien.

#### 4.1.3 Generation and Validation of Bike Flows

Finally, typical bike flows are generated providing 24 time-dependent OD matrices for all 59 stations. The OD matrices contain a total of 1569 daily trips performed with 627 bikes. The information model distributes these trips to 59 × 59 × 24 = 83544 OD pairs. By scaling and transformation, different demand scenarios can be generated. We use the original data set (1569 trips, 2.5 trips per bike) as the basic demand scenario and create a second scenario, the high demand scenario, with twice the demand (3138 trips, 5 trips per bike).

An extract from the resulting OD matrices of the basic demand scenario for a particular time period is shown below. Rows present origin stations, columns destination stations and matrix entries the bike flows. The left matrix shows real-valued bike flows as computed by the information model. In order to transform the real-valued bike flows into integer bike flows, a threshold of $\tau = 0.1488$ is determined for rounding such that the total number of rounded flows amounts to 1569 trips. Applying this threshold yields the integer bike flows shown in the right matrix. For instance, a value of $0.1488 > \tau$ translates into 1, and a value of $0.1386 < \tau$ translates into 0.

$$\begin{bmatrix} 0.1765 & 0.0967 & 0.2372 & ... \\ 0.1386 & 0.9095 & 0.1900 & ... \\ 0.3463 & 0.1550 & 0.4201 & ... \\ ... & ... & ... & ... \end{bmatrix} => \begin{bmatrix} 1 & 0 & 1 & ... \\ 0 & 1 & 1 & ... \\ 1 & 1 & 1 & ... \\ ... & ... & ... & ... \end{bmatrix}$$

In order to generate typical bike flows for the high demand scenario, real-valued bike flows are doubled and a new threshold $\tau = 0.2104$ is determined such that the total number of rounded flows amounts to 3138 trips. Note that not each integer bike flow of the basic demand scenario is doubled in the end, leading to the new set of integer bike flows as shown below:

$$\begin{bmatrix} 0.3530 & 0.1934 & 0.4744 & ... \\ 0.2772 & 1.8190 & 0.3800 & ... \\ 0.6920 & 0.3100 & 0.8402 & ... \\ ... & ... & ... & ... \end{bmatrix} => \begin{bmatrix} 1 & 0 & 1 & ... \\ 0 & 2 & 1 & ... \\ 1 & 1 & 1 & ... \\ ... & ... & ... & ... \end{bmatrix}$$

By applying this approach, we assume that it is more important to consider the variation of more significant flows in SND, while variation of inferior flows must be handled by operational planning. In the end, OD pairs with a flow of a very small expected number of bikes are considered as not relevant for SND, while significant bike flows are amplified.

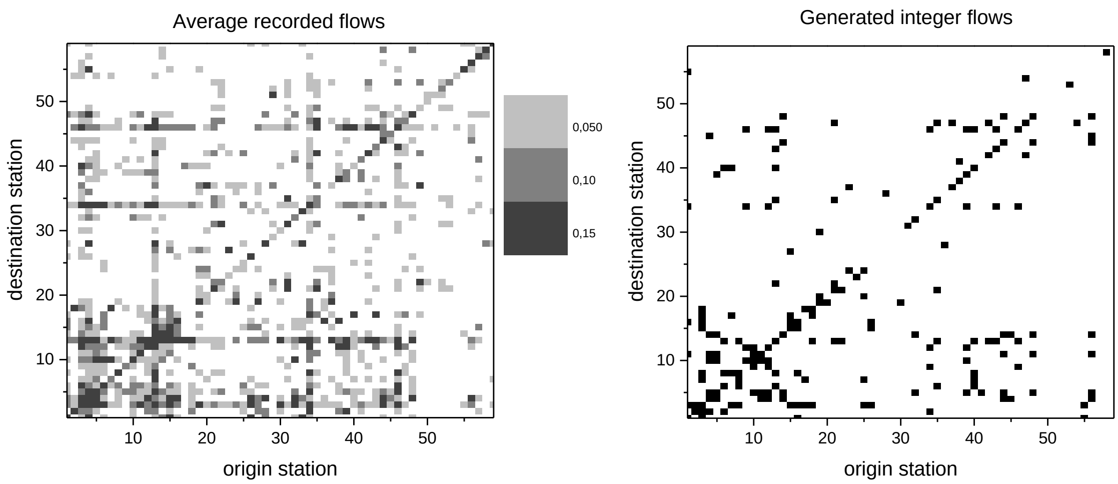


**Fig. 6** Comparison of the general flow structure of average recorded flows and generated integer bike flows

For validation, we explore the general flow structure as well as temporal and spatial characteristics of bike flows. To this end, we compare average bike flows as recorded by the BSS with generated integer bike flows as described above in contour diagrams (Fig. 6) for afternoon bike rentals. For all OD pairs, the diagrams show the intensity of the flows; the larger the flow, the darker the color. The average recorded bike flows stand out due to numerous OD pairs with a small value in the range of $[0; 0.65]$. Thus, simply averaging bike flows is not adequate for tactical planning. Due to the small sample size, the influence of non-typical behavior in trips at individual stations, e.g., due to sudden rain, events and full or empty stations, could superimpose the main trip purposes. Based on integer flow generation, the more significant bike flows translate into integer flows $\{0,1\}$ that are considered by SND as highlighted by the right diagram.

Focusing on temporal characteristics, Fig. 7 compares the rental activity of average recorded flows and generated integer bike flows aggregated by the hour of the day. In general, the shape of the curves are quite similar, both showing the night, morning and afternoon peaks. The differences in rental activity range between -2 and +5 percentage points. The rental activity in low usage hours is mitigated, whereas the rental activity in high usage hours is amplified by the information model. These characteristics follow our intuition of data aggregation for SND. We consider the differences between average recorded and generated integer bike flows to be within reasonable limits, carving out the typical temporal characteristics of Citybike Wien.

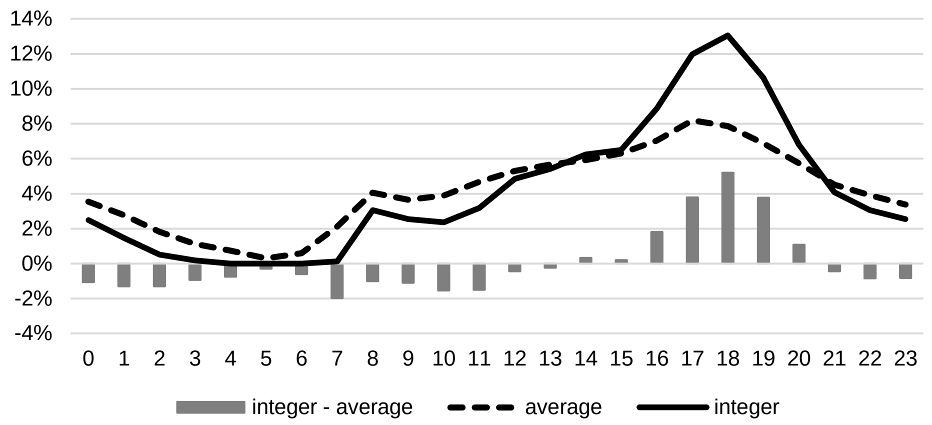


**Fig. 7** Comparison of average and generated integer bike flows per hour of the day

Investigating spatial characteristics of average recorded and generated integer bike flows for each station, Fig. 8 plots the total daily number of average rentals against generated integer rentals. The different colors denote the activity cluster the station was assigned to. For instance, the most upper right working cluster station has 76 average rentals and 108 integer rentals. Again, the number of rentals at stations with a high number of average recorded rentals is amplified by our information model (see cluster working stations, for example). In contrast, rentals at stations with a low number of average recorded rentals are mitigated (see mixed cluster stations, for example). As a result, for generated integer values, 40% of the stations account for 80% of rentals. For average values of recorded flows, 64% of the stations account for 80% of rentals. Thus, stations that are frequently used on average play a significant role in tactical planning, whereas rarely used stations are disregarded by our information model. Consequently, stations with typically high usage rate face higher relocation demand leading to presumed higher relocation demand. Similar findings are observed for returns.

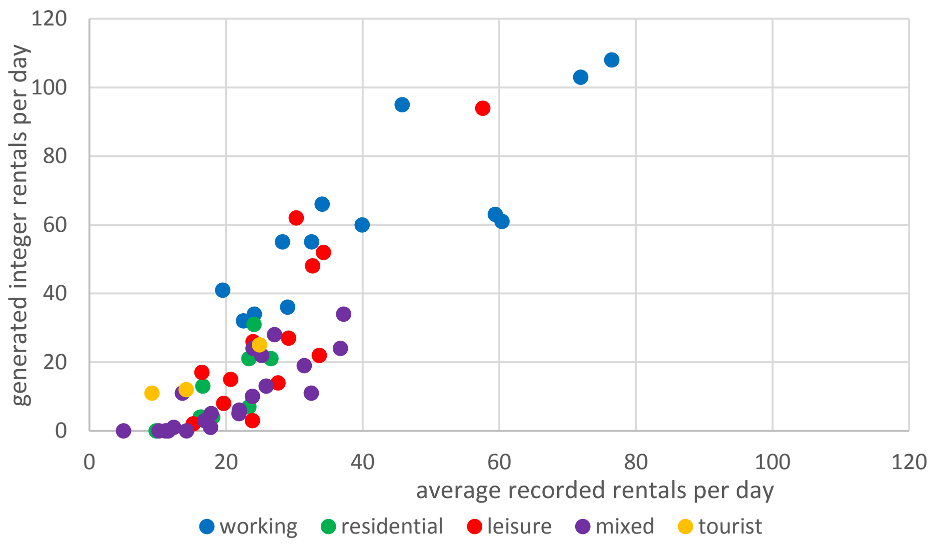


**Fig. 8** Average number of rentals compared to generated integer number of rentals per station

In sum, the information model carves out typical flows while retaining the spatio-temporal characteristics of Citybike Wien. It can be concluded that the information model reasonably generalizes the user behavior for tactical planning purposes. The generated integer bike flows provide an adequate input for SND.

### 4.2 Service Network Design for Different Demand Scenarios

We turn our attention to the application of SND for different scenarios of demand for Citybike Wien. We first describe the experimental setup with particular focus on the parameters and the computational solution environment (Sect. 4.2.1). Based on these parameters, we conduct SND for the basic demand scenario and compare the results to SND for the high demand scenario. The key figures of the solutions are compared in Sect. 4.2.2. The characteristics of fill levels and relocation services are discussed in Sect. 4.2.3 and Sect. 4.2.4, highlighting the extent that stations need relocation services to support operational planning.

4.2.1 Experimental Setup

The experimental setup for SND is as follows:

- Two demand scenarios: basic demand (1569 trips) and high demand (3135 trips)
- The network of Citybike Wien comprises $n = 59$ bike stations with a total number of 1253 bike racks and a total of $b = 627$ bikes (~50% average fill level).
- Time is discretized in terms of $t_{max} = 24$ (hourly) time periods.
- We assume that relocation services take one hour on average (approx. 15-20 minutes for loading and unloading plus travel times between stations).

- According to the system operator, handling costs depend on the time of the day. Daytime handling costs are set to $ch_{day} = 4$ Euro (in effect for time periods 8 to 17), while night time handling costs are more expensive ($ch_{night} = 7$ Euro).
- Transportation costs are assumed to be independent of the time of day and amount to $ct_{ij} = 0.5$ Euro per kilometer.
- The lot size of relocation services is $l = 20$.
- Bike and bike rack safety buffers are set to zero for each station and time period, ensuring that fill levels are non-negative and do not exceed station capacities.

The MIP model described in Section 3 is implemented in IBM ILOG OPL and solved with CPLEX 12.5 on an INTEL Core i5 processor at 3.2 GHz and 8 GB RAM running Windows 7 64 Bit. Both demand scenarios are given 30 minutes run time. CPLEX returns solutions with a gap of 1-2%. Although these gaps are very small, the optimal solution could still not be obtained after an increased run time of 24 hours. Note again that this instance, compared to other BSS, is a small instance with a total of 59 × 59 × 24 = 83544 binary relocation service variables. For bigger instances, a heuristic approach will be needed due to the sheer number of binary variables.

#### 4.2.2 Key Figures of the Service Network

First, we discuss key figures resulting from SND for the different demand scenarios. In order to demonstrate the benefit of optimized fill levels, we compare the costs of relocation required for the "optimal" fill levels to manually preset "naïve" fill levels. As often suggested by practitioners, we set the "naïve" fill levels for all stations to 50% in the hour of the lowest demand (hour 5). Table 1 summarizes these figures in terms of the number of relocated bikes, the number of relocation services, average number of relocated bikes per service as well as total and relative costs of relocation. The relative costs of relocation can be interpreted as the "usage fee" per trip required to compensate relocation costs.

**Table 1** Key figures of relocation services

| Demand scenario | Relocated bikes | Relocation services | Relocated bikes per service | Total relocation costs | Relative relocation costs |
|---|---|---|---|---|---|
| basic (naïve) | 130 | 42 | 3.09 | 584 | 0.37 |
| basic (optimal) | 119 | 32 | 3.71 | 496 | 0.32 |
| high (naïve) | 282 | 70 | 4.02 | 1345 | 0.42 |
| high (optimal) | 215 | 46 | 4.67 | 932 | 0.30 |

For the basic demand scenario, "naïve" fill levels result in 130 relocated bikes with 42 relocation services. Each relocation service carries 3.09 bikes on average. Total costs for relocation services amount to 584 Euros. A "usage fee" of 0.37 Euros per trip would thus be required to compensate relocation costs. In contrast to the "naïve" fill levels, "optimal" fill levels result in significantly lower relocation costs (17%), namely 496 Euros (119 relocated bikes with 32 relocation services). For the high demand scenario, the benefit of "optimal" fill levels becomes even more significant. “Naïve” fill levels result in 1345 Euros relocation cost (282 relocated bikes with 70 relocation services) whereas a saving of 44% is achieved with “optimal” fill levels. Comparing the results of “optimal” fill levels for the basic and high demand scenario shows that the number of relocated bikes increases by factor 1.8, the number of required relocation services by a factor of about 1.5 and the total relocation costs by a factor of 1.87. It is of note that doubling the demand does not result in doubled relocation services and costs. With doubled demand, the relative costs of relocation decrease slightly. Due to consolidation of relocation services, service capacities are utilized better, and only few additional relocation services are required. Furthermore, adapted fill levels compensate increased demand to a certain extent.

#### 4.2.3 Characteristics of Fill Levels

The aim of optimized fill levels is to ensure efficient provision of service at all times of the day. We present and evaluate the optimized fill levels for the two demand scenarios and the morning and afternoon peak hours. They are depicted in Fig. 7 by means of box plots of fill levels per cluster and hour of day.

The basic demand scenario reflects the demand for a typical working day in the summer season. In the morning peak hour, stations belonging to the working cluster require a low fill level of about 18% on average, and stations of the other clusters require a high fill level of about 60% and 70% on average. Thus, empty bike racks are needed at working cluster stations, whereas bikes are required at other cluster stations. In the afternoon peak hour, stations of the working cluster require higher fill levels than stations of the residential clusters. Fill levels at working cluster stations are almost 50% on average and almost 40% at residential cluster stations. Striking is the high variance of fill levels of the working and residential cluster compared to the morning peak. Fill levels at other clusters remain about the same. Note that the variance of fill levels among stations is high in general. For the working, residential and leisure cluster stations, capacity is sufficient to reserve bikes or bike racks for the demand of the upcoming time periods. Regarding mixed cluster stations, the high variance occurs due to diverse trip purposes. Tourist stations seem to serve as “buffer” stations being (almost) full or (almost) empty because the demand in general is rather low. In sum, fill levels reflect the rental and return activity of the clusters.

For the high demand scenario, average fill levels are more distinct and the variance within individual clusters is lower. Generally, the system seems to be more used to capacity which is reflected by the more distinct fill levels with smaller variance. In the morning peak hour, the higher demand induces more returns at working cluster stations and more rentals at residential cluster stations. Thus, more bike rack capacity is required at working cluster stations and more bike capacity is needed at residential cluster stations. As a result, fill levels at working cluster stations are 8% on average and 95% at residential cluster stations. In the afternoon peak hour, fill levels at stations of the working and residential clusters are more distinct than in the basic demand scenario for the same reason. Missing bike capacity is compensated by means of the leisure cluster stations, which show lower fill levels in the morning and afternoon.

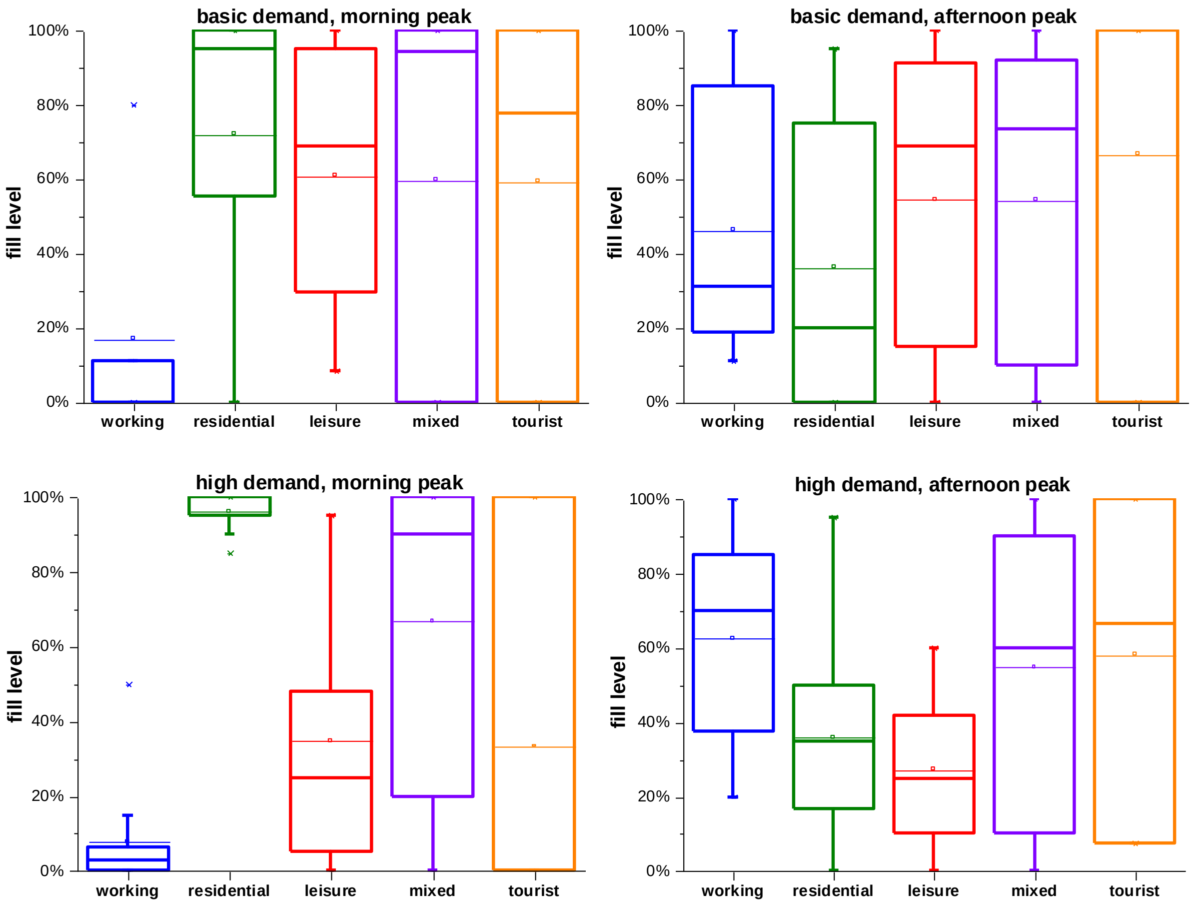


**Fig. 7** Boxplots of fill levels per cluster for peak hours in two demand scenarios

In sum, the demand increase especially affects working and residential cluster stations. Stations belonging to the leisure cluster serve as compensation stations for the increased demand for rentals and returns at working and residential cluster stations. Hence, these stations play a vital role in tactical planning.

#### 4.2.4 Characteristics of Relocation Services

Characteristics of relocation services determined by SND can aid the operator in preparing and implementing relocation services. We present spatio-temporal characteristics of relocation services resulting from SND for the basic and high demand scenarios.

Fig. 8 shows the total number of bikes that are expected to be picked up and returned by relocation services at each station, arranged by cluster assignment for the basic and high demand scenarios. We can clearly identify stations that require relocation pickups, returns or stations that can compensate demand without relocation. Regarding the basic demand scenario, relocation demand ranges between 20 relocation pickups and 14 relocation returns, i.e., the first station in the working cluster requires 14 bikes to be returned by relocation services, whereas the last station in the residential cluster needs 20 bikes to be picked up by relocation services. Stations that require relocation returns mainly belong to the working cluster, and stations requiring relocation pickups belong to the residential cluster. Stations of the leisure and mixed clusters need both relocation pickups and relocation returns, but they are also often able to (almost) balance pickups and returns properly without relocation services.

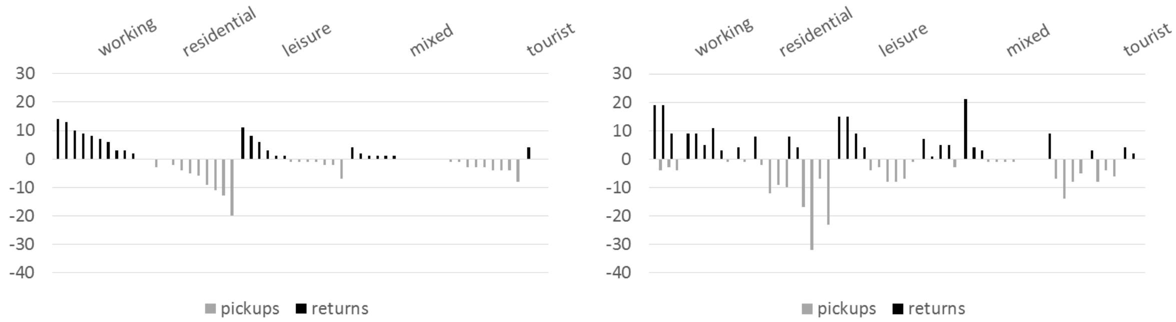


**Fig. 8** Total number of returned (positive) and picked up (negative) bikes by relocation for the basic demand scenario (left) and high demand scenario (right)

Regarding the high demand scenario, the presented order of stations is the same as in the low demand scenario. Higher demand causes increasing relocation demand, ranging from 32 relocation pickups to 21 relocation returns. Comparison of the two demand scenarios shows that the tendency of a station requiring either relocation returns, relocation pickups or no relocation remains for 66% of the stations when demand increases. At 7% of the stations, the required relocation efforts decrease. For 22% of the stations, the type of relocation service swaps from pickups to returns or vice versa.

Findings from SND show that stations either require relocation pickups or returns in general. This finding may support the planning of relocation operations, since the direction of relocation is known. Furthermore, SND gives indications on the priority of relocation operations at stations. Stations requiring a high number of relocation pickups or returns may be visited once a day. Stations with a medium number of relocation pickups or returns need relocation only on certain days of the week. The remaining stations may be visited occasionally. Important for operational and strategic planning is that there are three stations in the high demand scenario that require both relocation pickups and relocation returns. This implies insufficient capacity, because these stations cannot compensate demand variation throughout the day. Implications for the operational level are that these stations require relocation services more than once a day. Implications for the strategic level are that the size of the station should be extended, if possible.

Overall, the computational experiments show that SND helps determining reasonable fill levels and relocation services. The benefit of this tactical approach is that determined fill levels may serve as target fill levels for operational planning. Furthermore, characteristics of relocation services can aid the operator in the planning of relocation tours. SND provides information on the expected relocation demand at stations and shows which stations might play a crucial role in operations. Furthermore, information on the expected flows of relocations can help reducing the complexity of operational planning tasks such as routing of service vehicles.

## 5 Conclusions and future research

This paper lays the foundation of SND for BSS and related shared mobility systems. We have proposed an integrated approach of intelligent data analysis and mathematical optimization for SND in BSS. An information model has been presented allowing for the generation of spatio-temporal bike demand in terms of time-dependent OD matrices. Derived OD matrices have served as input for a MIP based SND model. The optimization model determines the optimal fill level at stations minimizing the expected costs of relocation services while ensuring a predefined service level. The computational study has shown that the approach yields reasonable fill levels at stations and identifies the extent that stations need relocation. The presented information model represents dynamic system behavior as required by SND. Note that it could also be used to depict typical bike flows for strategic planning and individual trips for operational planning if parameterized accordingly. For strategic planning, spatio-temporal distributions of not yet implemented stations could be forecasted from existing stations, e.g., one could assume that a new station in a residential area will likely show similar main trip purposes as an existing station in a similar area. For operational planning, the expected values would have to be incorporated into probability functions such as the Poisson distribution. In combination with the distribution of trip durations, this would allow for generation of individual trips between stations. Finally, the information model is also applicable to other shared mobility systems as long as a sufficient amount of operational data is available. Future research could investigate improved ways of modeling relocation services for BSS and the adaption of the SND model to related shared mobility systems. We want to encourage future research regarding SND models better anticipating operational relocation decisions, for example, by modeling routing of relocation vehicles. Furthermore, a more robust service network design would be desirable. Generated demand scenarios reflect differences in demand variation. These scenarios could be applied to a stochastic optimization model. Regarding the evaluation of tactical decisions, it might be interesting to study the influence of fill levels on resulting relocation tours. Another interesting future path is the determination of optimal safety buffers for each station and period to improve target fill levels. Whilst the above information model is generally applicable to shared mobility systems, the SND model has to be modified for other means of transportation, system designs and business models.

Acknowledgements

We thank „Gewista Werbegesellschaft m.b.H.“ for providing data from their project “Citybike Wien” and in particular Dr. Hans-Erich Dechant for the helpful discussion of related issues and results.

## References

Benchimol M, Benchimol P, Chappert B, et al. (2011) Balancing the stations of a self-service bike hire system. RAIRO-Operations Research 45:37–61.

Berthold MR, Borgelt C, Höppner F, Klawonn F (2010) Guide to intelligent data analysis. Springer

Bookbinder JH, Sethi SP (1980) The dynamic transportation problem: A survey. Naval Research Logistics Quarterly 27:65–87.

Borgnat P, Robardet C, Rouquier JB, et al. (2010) Shared bicycles in a city: a signal processing and data analysis perspective. Advances in Complex Systems (ACS) 14:415–438.

Boyaci B, Zografos KG, Geroliminis N (2015) An optimization framework for the development of efficient one-way car-sharing systems. European Journal of Operational Research 240:718–733.

Büttner J, Petersen T (2011) Optimising Bike Sharing in European Cities - A Handbook.

Caggiani L, Ottomanelli M (2012) A modular soft computing based method for vehicles repositioning in bike-sharing systems. Procedia-Social and Behavioral Sciences 54:675–684.

Campbell A, Clarke L, Kleywegt A, Savelsbergh M (1998) The inventory routing problem. Fleet management and logistics. Springer, pp 95–113

Cepolina EM, Farina A (2012) A new shared vehicle system for urban areas. Transportation Research Part C: Emerging Technologies 21:230–243.

Chow JY, Sayarshad HR (2014) Symbiotic network design strategies in the presence of coexisting transportation networks. Transportation Research Part B: Methodological 62:13–34.

Contardo C, Morency C, Rousseau LM (2012) Balancing a dynamic public bike-sharing system.

Correia GHA, Antunes AP (2012) Optimization approach to depot location and trip selection in one-way carsharing systems. Transportation Research Part E: Logistics and Transportation Review 48:233 – 247.

Crainic TG (2000) Service network design in freight transportation. European Journal of Operational Research 122:272–288.

Dell'Amico M, Hadjicostantinou E, Iori M, Novellani S (2013) The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. Omega

DeMaio P (2009) Bike-sharing: History, impacts, models of provision, and future. Journal of Public Transportation 12:41–56.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B (Methodological) 1–38.

Faghih-Imani A, Eluru N, El-Geneidy AM, et al. (2014) How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography

Di Febbraro A, Sacco N, Saeednia M (2012) One-Way Carsharing: Solving the Relocation Problem. Transportation research record 113–120.

Flyvbjerg B, Skamris Holm MK, Buhl SL (2006) Inaccuracy in traffic forecasts. Transport Reviews 26:1–24.

Fricker C, Gast N (2014) Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. EURO Journal on Transportation and Logistics 1–31.

Froehlich J, Neumann J, Oliver N (2009) Sensing and predicting the pulse of the city through shared bicycling. International Joint Conference on Artificial Intelligence. pp 20–26

Garcia-Palomares JC, Gutiérrez J, Latorre M (2012) Optimizing the location of stations in bike-sharing programs: a GIS approach. Applied Geography 35:235–246.

Di Gaspero L, Rendl A, Urli T (2013a) A hybrid ACO+CP for balancing bicycle sharing systems. Hybrid Metaheuristics. Springer, pp 198–212

Di Gaspero L, Rendl A, Urli T (2013b) Constraint-based approaches for balancing bike sharing systems. Principles and Practice of Constraint Programming. pp 758–773

George DK, Xia CH (2011) Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. European Journal of Operational Research 211:198–207.

Ho SC, Szeto W (2014) Solving a static repositioning problem in bike-sharing systems using iterated tabu search. Transportation Research Part E: Logistics and Transportation Review 69:180–198.

Johnston RA (2004) The Urban Transportation Planning Process. The geography of urban transportation

Jorge D, Correia G, Barnhart C (2012) Testing the validity of the MIP approach for locating carsharing stations in one-way systems. Procedia-Social and Behavioral Sciences 54:138–148.

Kaltenbrunner A, Meza R, Grivolla J, et al. (2010) Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. Pervasive and Mobile Computing 6:455–466.

Kaspi M, Raviv T, Tzur M (2014) Parking reservation policies in one-way vehicle sharing systems. Transportation Research Part B: Methodological 62:35–50.

Kek AGH, Cheu RL, Meng Q, Fung CH (2009) A decision support system for vehicle relocation operations in carsharing systems. Transportation Research Part E: Logistics and Transportation Review 45:149–158.

Kloimüllner C, Papazek P, Hu B, Raidl GR (2014) Balancing Bicycle Sharing Systems: An Approach for the Dynamic Case⋆.

Lin JR, Yang TH (2011) Strategic design of public bicycle sharing systems with service level constraints. Transportation Research Part E: Logistics and Transportation Review 47:284–294.

Lin J-R, Yang T-H, Chang Y-C (2013) A hub location inventory model for bicycle sharing system design: Formulation and solution. Computers & Industrial Engineering 65:77–86.

Martinez LM, Caetano L, Eiró T, Cruz F (2012) An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon. Procedia-Social and Behavioral Sciences 54:513–524.

Midgley P (2011) Bicycle-sharing schemes: enhancing sustainable mobility in urban areas.

Nair R, Miller-Hooks E (2011) Fleet management for vehicle sharing operations. Transportation Science 45:524–540.

Nair R, Miller-Hooks E (2014) Equilibrium network design of shared-vehicle systems. European Journal of Operational Research 235:47–61.

Nourinejad M, Roorda MJ (2014) A dynamic carsharing decision support system. Transportation Research Part E: Logistics and Transportation Review 66:36–50.

O’Brien O, Cheshire J, Batty M (2013) Mining bicycle sharing data for generating insights into sustainable transport systems. Journal of Transport Geography

Raidl GR, Hu B, Rainer-Harbach M, Papazek P (2013) Balancing bicycle sharing systems: Improving a VNS by efficiently determining optimal loading operations. Hybrid Metaheuristics. Springer, pp 130–143

Rainer-Harbach M, Papazek P, Hu B, Raidl GR (2013) Balancing bicycle sharing systems: A variable neighborhood search approach. Evolutionary Computation in Combinatorial Optimization. Springer, pp 121–132

Raviv T, Kolka O (2013) Optimal inventory management of a bike-sharing station. IIE Transactions 45:1077–1093.

Raviv T, Tzur M, Forma IA (2013) Static repositioning in a bike-sharing system: models and solution approaches. EURO Journal on Transportation and Logistics 2:187–229.

Ricker V, Meisel S, Mattfeld DC (2012) Optimierung von stationsbasierten Bike-Sharing-Systemen. Proceedings of MKWI 2012. pp 215 – 226

Sayarshad H, Tavassoli S, Zhao F (2012) A multi-periodic optimization formulation for bike planning and bike utilization. Applied Mathematical Modelling 36:4944 – 4951.

Schuijbroek J, Hampshire R, van Hoeve W-J (2013) Inventory Rebalancing and Vehicle Routing in Bike Sharing Systems.

Shaheen SA, Guzman S, Zhang H (2010) Bikesharing in Europe, the Americas, and Asia. Transportation Research Record: Journal of the Transportation Research Board 2143:159–167.

Shu J, Chou MC, Liu Q, et al. (2013) Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. Operations Research 61:1346–1359.

Vogel P, Greiser T, Mattfeld DC (2011) Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. Procedia-Social and Behavioral Sciences 20:514–523.

Vogel P, Mattfeld DC (2011) Strategic and Operational Planning of Bike-Sharing Systems by Data Mining – A Case Study. In: Böse, J and Hu, H and Jahn, C and Shi, X and Stahlbock, R and Voß, S (ed) Computational Logistics. Springer Berlin / Heidelberg, Germany, pp 127–141

Wang X, Lindsey G, Schoner JE, Harrison A (2012) Modeling bike share station activity: the effects of nearby business and jobs on trips to and from stations. Transportation Research Record 43:45.

Weikl S, Bogenberger K (2012) Relocation strategies and algorithms for free-floating Car Sharing Systems. Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on. pp 355–360

Zahory MN (2009) Request for Proposals for the operation of the Arlington Bike-sharing Program. http://www.metrobike.net/index.php?s=file_download&id=18. Accessed 30 Jun 2012