QUANTICOL

A Quantitative Approach to Management and Design of Collective and Adaptive Behaviours

quanti**col**

http://www.quanticol.eu

TR-QC-03-2015

Model-based Assessment of Aspects of User-satisfaction in Bicycle Sharing Systems

Extended Version

Revision: 1.0; May 27, 2015

Author(s): Mieke Massink (CNR), Rytis Paškauskas (CNR)

Publication date: May 27, 2015

Part. no.	Participant organisation name	Acronym	Country
1 (Coord.)	University of Edinburgh	UEDIN	UK
2	Consiglio Nazionale delle Ricerche – Istituto di Scienza e Tecnologie della Informazione "A. Faedo"	CNR	Italy
3	Ludwig-Maximilians-Universität München	LMU	Germany
4	Ecole Polytechnique Fédérale de Lausanne	EPFL	Switzerland
5	IMT Lucca	IMT	Italy
6	University of Southampton	SOTON	UK

COOPERATION

Contents

1 Introduction

Smart bicycle sharing is a form of public transport that provides short-term self-service bicycle hiring [Mai09, Mid11]. It has evolved a long way from the early ideas dating back to the sixties. Today, hundreds of cities worldwide have such programs, operating up to tens of thousands of vehicles and thousands of docking stations (e.g. Hangzhou or Paris[1]). Recent popularity of bike-sharing gained momentum with the introduction of information and smart card technologies, which improved service predictability, reduced the risks of theft or damage, and streamlined the subscription procedures. In those cities, smart bike-sharing has become a reliable mode of public transport, welcomed by the general public for its dependability and bicycle's environmental, societal, and health benefits [PDH10]. However, smart bike-sharing programs raise multiple issues concerning their carbon footprint [FWH14], integration with other modes of public transport, choosing proper service features [tBFG14], and understanding the effects of user incentives [FG14], to mention a few.

---

[1]Institute for Transportation and Development Policy of China, `http://www.publicbike.net/defaulten.aspx`; Vélib, `http://www.velib.paris.fr`
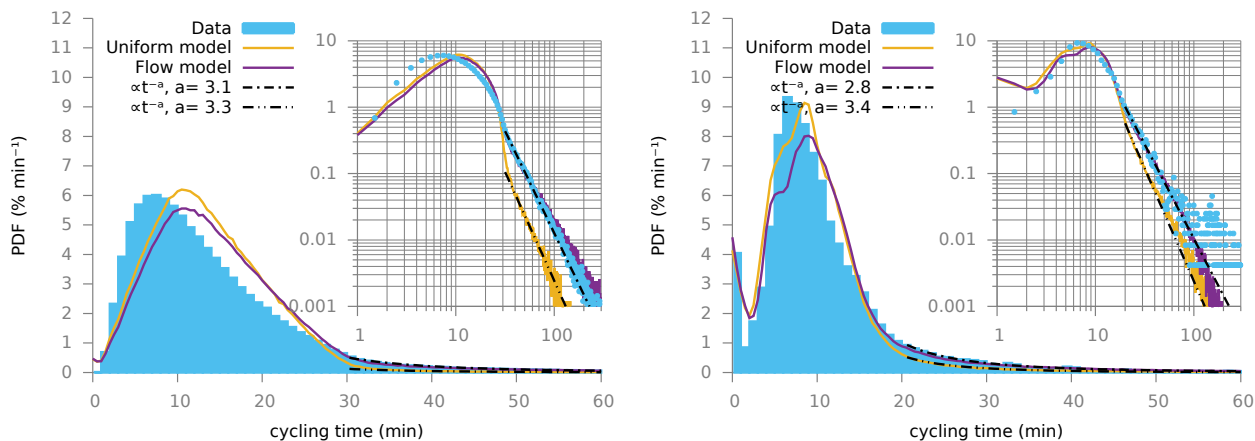
Figure 1: PDFs of cycling times (Data) in London (right) and Pisa (left) aggregated over one month (October 2013 and October 2014, respectively) and the predictions using a uniform model (light lines) and a flow model (dark lines).

In developed urban environments, the question that potential users of any public form of transport will be asking themselves increasingly often is 'which' mode of public transport to rely on rather than 'if' they will use public transport. This concerns smart bike-sharing too, since the majority of cycling trips in cities could also be made by a combination of walking and other modes of public transport, or by a private bike. A potential user may favour safety, health and environmental impacts of using a bike-sharing service, but there are likely to be other factors that favour some alternative option. The successful running of multiple public transport services may in the long term be determined not only by proper top-down planning, but also by the cumulative effect of 'micro-decisions' by the public, as the example of the bike-sharing system in Melbourne [Car14] suggests. Being able to evaluate the balance between services and policies could in the long run determine the success of some programs.

It is notoriously difficult to evaluate user satisfaction from the available data collected by the system. Typical bike-sharing data consists of static parameters of stations and fluctuating numbers of parked vehicles, either provided by the operator, or collected from a public domain. Oftentimes vehicle identification numbers are also available. They allow to relate hiring with the corresponding returning events and to visualise the dominant spatial and temporal vehicle flows in cities [FNO09, BAF+11, OCB14]. Naturally, this data concerns only such trips which actually, and thus successfully took place, and raises the issue of missing information about users who chose an alternative transport service. Moreover, even the successful trip data concerns only the middle part of the 'walk-cycle-walk' travel cycle. The missing links conceal the trip–objective relation, which is important to the evaluation of a system from the service efficiency perspective. Alternative approaches such as [BMLEG12] provide useful complementary insights, but they are susceptible to similar bias issues.

The main contribution of this paper is to show that a *model based approach*, that takes into account certain minimal assumptions about the user behaviour, can provide complementary insights into the performance of bike-sharing from a users' perspective. This is illustrated by showing that the aggregated cycling time distributions of real bike-sharing systems can be reproduced to a degree without the parameter fitting of real systems, or the use of privacy-sensitive user information. Furthermore, enriching the model in a step-wise manner suggests other generic insights into the multifaceted question of user satisfaction. We explore the interpretation of cycling trip durations in a manner akin to that of an 'actuary'. Statistics of human life's duration follow a certain probability law.[2] The subtle features of the probability density function (PDF), especially in the so-called 'tail' of the distribution, are oftentimes of most interest, since they have important consequences to the conditional expected

---

[2]Cf. the Gompertz-Makeham model.

life-times, and to the assessment of long term risks. In this article we will compare two data-sets, each consisting of cycling trips, aggregated over one month (October) and referring to bike-sharing systems in two cities: London (UK) and Pisa (Italy). As is evident from Fig. 1 (Data), the cycling time PDFs share remarkable similarities, such as a mode at about 10 minutes, or an algebraic 'tail' of the distribution beyond 30 minutes (cf. also Fig. 3 in Ref. [BAF$^+$11]). The latter part of the PDF is well approximated by $f(t) \propto t^{-1-a}$, with an exponent $a > 0$. Validity of this law spans nearly three decades and appears to be limited only by the sample size. In both cases, it predicts cycling trips whose duration exceeds the time necessary to traverse a corresponding city. One of the main motivating ideas of this article is to uncover a 'story' told by these cycling time PDFs. To capture the domain independence of such distributions, we will use an agent-based model with a basic assumption, that bike-sharing users' main concern is to save travel time and to arrive at the planned destination with a high probability within the expected time. We will show that their algebraic parts reveal more about the user satisfaction than any other characteristic does. The 'risks' and 'expected conditional life-times' will be related to the personal contingencies of being late at one's appointment. Varying the agent and station distributions, we will obtain the 'uniform', and the 'flow' models (whose PDFs are also shown in Fig. 1) and relate several differences of these models to the factors affecting user satisfaction with the system. This will be achieved using a simulation approach, applied to Markov Renewal Process models (MRP) [Ç75, Kul95]. Markov Renewal Process is the simplest stochastic modelling framework which can accomodate sufficient behavioural complexity required for our purposes, thanks to a possibility to use non-Markovian renewal PDFs.

The outline of the paper is as follows. Section 2 briefly recalls the essence of MRPs, provides a description of the bike-sharing model, including its justification, and describes the simulation method. The results concerning trip duration distributions are presented in Section 3, whereas models for Pisa and London are discussed in Section 4. Section 5 concludes with some further considerations and open issues.

2   The Bike-sharing Model

2.1   Markov Renewal Processes

Markov Renewal Processes (MRP) are a generalisation of Continuous Time Markov Chains to non-Markovian events, and non-exponential distributions of inter-event times [Ç75, Kul95]. In this section we briefly recall MRPs, and motivate their use for the modelling of bike-sharing user behaviour.

Let $(X, T) = \{X_i, T_i; i \in \mathbb{N}\}$ be a stochastic process taking place in $E^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}}$, where E is some countable set, representing the 'state space', and $\mathbb{R}_+ = [0, \infty)$ represents the time-line of evolution. Markov Renewal Process is a Kolmogorov model with the conditional probability given by

$$\Pr\{X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i\} = Q_{ij}(t), \tag{1}$$

for each pair of states $i, j \in E$, where $Q_{ij}(t)$ is a right-continuous, non-decreasing and bounded function satisfying $Q_{ij}(\infty) \leq 1$ (a so-called càdlàg function) and $\sum_j Q_{ij}(\infty) = 1$. A matrix $Q = (Q_{ij}(t); i, j \in E)$ with these properties is called a semi-Markov kernel of (X, T). It is easily shown that a matrix $P = (P_{ij})$, whose elements are defined by $P_{ij} = Q_{ij}(\infty)$, is a stochastic matrix, and that functions $F_{ij}(t) = Q_{ij}(t)/P_{ij}$, for each $i, j \in E$, are distributions. As a consequence, $X = (X_n; n \in \mathbb{N})$ is a Markov chain (DTMC) with state space E and transition matrix P, i.e.

$$\Pr\{X_{n+1} = j \mid X_n = i\} = P_{ij}, \tag{2}$$

and the distribution of *sojourn time* in a state i, conditional on a subsequent jump to a state j, is given by $F_{ij}(t)$:

$$\Pr\{T_{n+1} - T_n < t \mid X_n = i, X_{n+1} = j\} = F_{ij}(t). \tag{3}$$
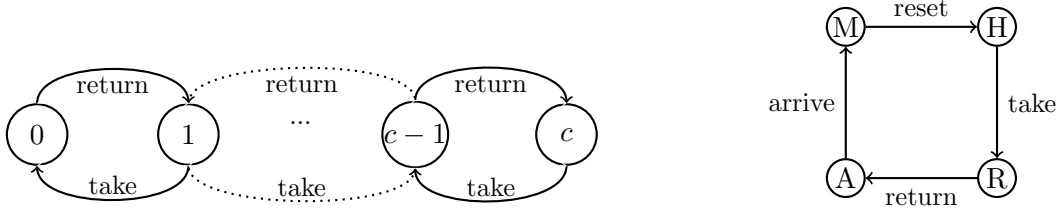
Figure 2:   The automata of a bicycle station (left) and a user-agent (right). The state space corresponding to a system of A agents and S stations, is a cartesian product $E = \prod_{i=1}^{S}\{0, \ldots, c_i\} \times \{H, R, A, M\}^A$.

A large group of probabilistic models used in tranport modelling can be interpreted as MRPs with a special choice of the kernel, most notably with kernels that are restricted to functions of the current state only. Examples include continuous time Markov chains (CTMC) and various queueing models [Ç75].

2.2   Motivation for the use of MRP

The MRP generalises a Markov Process in two aspects: it provides a mechanism to use arbitrary distributions (and not only exponential ones), and it allows to use transitions, conditioned on a current state *and* on the state to be entered subsequently. These are the main features used in what follows.

We will assume that bike-sharing users are time-conscious people whose decision to use bike-sharing is determined by the concern to save travel time, and to reach their objective at the expected time with high degree of certainty. If we accept this premise then we must also accept that the speed of travel is a major factor in the competitiveness of various modes of transport. To take the speed of travel into account in a stochastic model, it is easy to show that transition rates must be functions of both the current and future states, and that the probability distributions are not exponential. Let the state space represent an 'address book' of all the stations; we may take $E = \{1, \ldots, S\}$ where S is the number of stations, and each index is uniquely associated to some address $\mathbf{x}_i$. For an arbitrary pair of indices $i \neq j$, consider a trip from $\mathbf{x}_i$ to $\mathbf{x}_j$ along a fixed path, traversed at a constant pace p, measured in minutes per kilometer. The duration of this trip is $T = p \, |\mathbf{x}_i - \mathbf{x}_j|$, where $|\cdot|$ is the length of the path in kilometers. This result can be given a distribution function $F_{ij}(t) = 1 - 1_{t \leq t_{ij}^a}$, where 1_A is an indicator function, equal to one if A is true and zero otherwise, and $t_{ij}^a = p \, |\mathbf{x}_i - \mathbf{x}_j|$ is the so-called *activation time*. Clearly, $F_{ij}(t)$ is not an exponential distribution for any pair of indices, and its parameter t_{ij}^a is a function of the current state i, and a possible future state j. This argument is easily generalised to stochastic travel processes. Any trip between a pair of distant locations (i, j) will take a human traveller at least some finite time $t_{ij}^a > 0$, so that $F_{ij}(t) = 0$ if $t \leq t_{ij}^a$ [3]. The exponential distribution, on the other hand, is characterised by $t_{ij}^a = 0$ thus it allows arbitrarily fast travelling.

2.3   An outline of the model

The bike-sharing model is a generalisation of these motivating ideas. It describes a population of agents and bicycle stations. A two-dimensional rectangle is used to represent a city. The population of agents and the array of stations are contained within this area.

---

[3] A bound of t_{ij}^a, obtained by substituting the speed of light for $1/p$ gives absolute certainty, but also much larger bounds, assuming much slower speeds, can be used with near certainty.
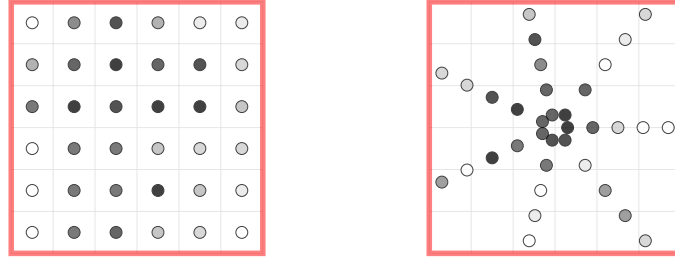
Figure 3: Square 6×6 and star-shaped 5×7 station configurations in a 3×3 km square.

2.3.1 Population of bicycle stations

A station is represented by a triple $(n, c, \mathbf{x})$, where n is the number of available bicycles (the *occupation number*), c is the *capacity*, and $\mathbf{x}$ is the geographical coordinate (the *address*) of a given station. An automaton of a typical station with capacity c is shown in Fig. 2. The configuration of stations can be arbitrary. Two different configurations, one rectangular and one star-shaped, are shown in Fig. 3. The total number of bikes is $N = \sum_{i=1}^{S} n_i$, and the total capacity $C = \sum_{i=1}^{S} c_i$. Fixed capacity and instantaneous transaction approximations are assumed throughout.

2.3.2 Population of user-agents

A user-agent combines several human factors pertaining to travelling and decisions. Each agent is parameterised by two addresses that specify the agent's origin and destination locations. The cycling and walking paces, and corresponding rates, are considered as random numbers sampled from a normal distribution. The typical human speeds of 5 km/h for walking, and 12 km/h for cycling, yield paces of 12 min/km, and 5 min/km, respectively, whereas the mean of both rates is set to one (min^{-1}). The agent states (see Fig. 2) are denoted and interpreted as follows: H= {wants to hire a bike}, R= {wants to return a bike}, A= {wants to arrive}, and M= {wants to reset}. A single 'walk-cycle-walk' travel cycle is quantified by a sequence of transitions H→R→A→M, with transition epochs T_1, T_2, T_3, T_4. The total duration of a trip is $T_4 - T_1$, whereas the duration of its cycling part is $T_3 - T_2$. An additional 'mutation' transition M→H is added to make agents' life cyclic and states recurrent, and allow continual regeneration of their objectives.

2.3.3 Stochastic dynamics

Agents drive the system by spontaneous decisions. There are two types of agent decisions that result in *firing* or *mutation* transitions. Firing transitions are further distinguished as either 'take' or 'return' transitions. They are synchronised in an obvious manner with two kinds of state-changes occurring at bicycle stations (see Fig. 2). The remaining 'arrive' and 'reset' transitions are *mutation* transitions. They are defined by *not* being synchronised with the station updates. Both re-initialise the agent states, the first one resulting in the arrival at a destination, the second one in a complete regeneration of its objectives. Agents' objectives can be initialised in various ways. In this article, we consider the so-called *spot commuter* mutation protocol. A spot commuter selects a random new pair of locations and the time until activation, both sampled from appropriate distributions. We remark that other protocols can be easily designed where, for example, the arrival epoch, rather than the departure moment, is a relevant issue. It is important to emphasize that agents can estimate the expected arrival epoch using a markovian forecasting protocol and, consequently, measure the late arrival epoch as a difference between the expected and actual arrivals.

The model that is used for the travel process is a composition of a *conditional travel process* and a *station utility model*, addressing two major sources of uncertainty of travel in urban environments. The
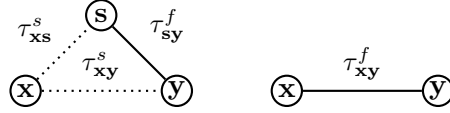
Figure 4:   Spatial (left) and temporal (right) decision criteria of participation.

conditional travel process addresses the randomness due to various interactions with the environment. It states that, conditional to fixed end points and a fixed itinerary, the travel time is a stochastic variable that corresponds to a first passage process of a one-dimensional random walk. The so-called renewal distribution of this process is the inverse Gaussian distribution [Chh88] which is further approximated by a delayed exponential distribution function $F(t) = 1 - 1_{t \le t^a} e^{-r^a(t-t^a)}$. Here, $t^a = p^a d$ is the activation time of an agent a, p^a is its travel pace, d is the distance, and r^a is as arrival rate, related to the diffusion property of a random walk. The station utility model, combined with the assumption of stochastic dynamics, address the decision process under uncertainty. Following the von Neumann-Morgernstern axiomatic approach to the description of such decisions, the existence of *station utility functions* with respect to hiring and returning is posited in the form $u_{ij}(\tau)$, for $i, j \in E$, with a control parameter τ, called the *decision scale* parameter. It is a 'motivational' parameter, describing the perceived utility of a station from an agent's point of view, and influences the agents' decision to return a bike to a particular station. A model for station utility perception is proposed in which the probability of $\{d > x\}$ for large distances x is given by, approximately, $\Pr\{d > x\} \sim \exp -\frac{p_f^2 x^2}{2\tau^2}$. Thus, agents with a larger τ value tend to search for suitable stations in a greater area surrounding their target. However, venturing further away from the destination increases the walking fraction of the trip, so that the total trip duration is likely to increase. On the other hand, smaller values should lead to shorter trips, provided that a suitable station can actually be found within the search area.

Composition of the two models yields a renewal function $F_{ij}^a(t)$ for an agent's a arrival epoch T at a station j, given the current state i, as

$$\Pr\{T \le t\} = F_{ij}^a(t) = 1 - 1_{t \le t_{ij}^a} e^{-u_{ij}(\tau^a) r^a(t-t_{ij}^a)}. \tag{4}$$

2.3.4   Elective participation and posterior evaluation

Agents are provided with the capacity to decide whether to accept or reject bike-sharing as a means of achieving their objective, and to measure the effectiveness of a trip in the case of acceptance. They decide whether a bike-sharing trip is a viable alternative to walking by estimating a kind of triangle inequality. Assuming that agents know the distance from the origin ($\mathbf{x}$) to the destination ($\mathbf{y}$) and to the neigbouring stations ($\mathbf{s}$) in advance, and their physical parameters, they estimate the expected travel time using a station $\mathbf{s}$ as $\tau_{\mathbf{xs}}^s + \tau_{\mathbf{sy}}^f$, and compare it against the estimated time of walking directly to the destination, $\tau_{\mathbf{xy}}^s$ ('s' and 'f' refer to walking and cycling, respectively, see Fig. 4). A station is accepted as a candidate for a trip if it satisfies the *triangle inequality*

$$\tau_{\mathbf{xs}}^s + \tau_{\mathbf{sy}}^f < \tau_{\mathbf{xy}}^s. \tag{5}$$

If at least one station in the network satisfies (5), a bike-sharing trip is accepted, otherwise it is rejected.

Agents estimate the cycling time from the origin to destination, $\tau_{\mathbf{xy}}^f$. A trip is accepted only if

$$\tau_{\mathbf{xy}}^f < t_c, \tag{6}$$
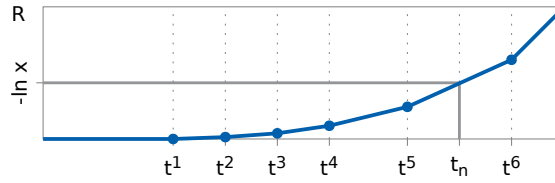
where t_c is the *cycling tolerance* parameter.

Figure 5:   The first half of the stochastic simulation algorithm determines the sojourn time (t_n) and the participating channels in the next transition (M_n).

For an acceptable trip, agents estimate the efficiency of a trip by comparing the actual trip duration $T_{\text{tot}} = T_4 - T_1$ with the estimated trip duration, had the agent walked the same itinerary, i.e. with $\tau^s = \tau_1^s + \tau_2^s + \tau_3^s$ (where indices 1,2,3 denote the consecutive parts of the 'walk-cycle-walk' travel cycle). To decide whether using bike-sharing has been a winning strategy, agents estimate the *efficiency ratio*

$$e = \frac{\tau^s}{T_{\text{tot}}} \, . \tag{7}$$

Thus agents can decide, retrospectively, whether bike-sharing saved time $(e > 1)$ or if it was a waste of time $(e \leq 1)$. Inclusion of the marginal case as failure is motivated by the assumption that walking is preferred to cycling by default.

2.3.5   Relation to Collective Adaptive Systems

The proposed bike-sharing model describes a collective system in a sense of a collective of agents that act concurrently. These agents do not interact directly. They represent non-cooperating entities with independent objectives. However the decisions that they make directly affect the stations. The states of stations influence other agents' decisions, so that decisions of one agent influences the decisions of all agents indirectly through the stations, and may be interpreted as a kind of weak interaction between agents. This approach is appropriate to real bike-sharing systems, if we adopt the common view that the participants go about their objectives by and large independently of what others' objectives are.[4] This may well be only a 'first order' approximation, but we claim it is a reasonable one. We will attempt to substatiante this claim by providing numerical evidence that the model yields meaningful predictions in section 3, that it captures the essential features of data from real bike-sharing systems and, moreover, yields fruitful interpretation of the user experience that could be tested, in section 4. We do not attempt to fit the model to the the 'invisible hand', or information paradigms of collective adaptive systems (CAS) of behavioural economics[5]. However, the bike-sharing model is consistent with a kind of 'weak' form of adaptivity, because a limited form of agent congnition is taken into account. It is incorporated into the Kolmogorov and the station utility models and is built on some very basic input from the two system approach to decision-making of people [Kah03]. The cognition and spontaneous decisions are embodied by, respectively, a loose survival policy in the guise of time-saving assumption which, in its turn, is a consequence of the causality of stochastic travel processes, and on the assumption that spontaneous decisions can be considered as random events that are conditionally independent on the past. As a next order approximation, modifications to the utility model could be considered to take into account variations of human decision making based on the instantaneous environmental conditions, and other forms of adaptation by introducing user incentives, designed with the aim to help redistribute bicycles from mostly full to mostly empty stations.

---

[4]assuming that the collision avoidance among agents and similar manouvering is irrelevant, at least to the first order in approximation

[5]see e.g. [MP09] for an introduction to CAS in behavioural economics

2.4   Method of simulation

Discrete stochastic processes are oftentimes simulated using methods that produce statistically exact sample paths [Gil76, And07]. The basic method requires two random number generations per step, one for each member of the pair (X, T). Thus, a statistically exact 'first reaction' of an M-channel Markov Process with rates $\lambda_1(X_n)$, ..., $\lambda_M(X_n)$ at the nth step could be determined as follows: The *sojourn time* $t_n = T_{n+1} - T_n$ in a current state X_n is drawn from an exponential distribution $1 - \exp(-\lambda(X_n)t)$, where $\lambda(X_n) = \sum_{i=1}^M \lambda_i(X_n)$, whereas the *next transition channel m* is drawn from a discrete M-point distribution with weights $\hat{\lambda}_i(X_n)$, $i = 1, \ldots, M$, where $\hat{\lambda}_i = \lambda_i/\lambda$.[6] A practical algorithm consists of drawing two random numbers $x, y \sim U(0, 1)$ from a uniform distribution U, then letting $t_n = -\frac{1}{\lambda(X_n)} \ln x$, and letting m be such that $\hat{\lambda}_1 \leq \cdots \leq \sum_{i=1}^m \hat{\lambda}_i \leq y \leq \sum_{i=1}^{m+1} \hat{\lambda}_i \leq 1$[Gil76, And07, GHP13]. Alternative statistically exact methods [GB00], and other, exact and approximate variations exist to address specific issues viz. multiple time-scales (see, e.g. [GHP13] and references therein).

Although the bike-sharing model is non-Markovian (see section 2.2), the same idea of simulating Markovian Processes can be adapted to obtain a statistically exact simulation of the bike-sharing model. In this case, each agent-station pair must be considered as a possible transition channel (thus M can be quite large) and the total transition rate in a state X is obtained as a sum of rates from all channels, the rate for each channel being $r_{ij}^a(t; \tau^a) = 1_{t>t_{ij}^a} u_{ij}(\tau^a)p^a$. Note that although a one agent – one station pair distribution is (a delayed) exponential, the system of many stations or many agents gives a non-exponential distribution of the form $1 - e^{-R(t)}$. The phase, $R(t)$, defined as an integral over the total rate, $R(t) = \int_0^t \sum r_{ij}^a(t'; \tau^a)dt'$, is illustrated in Fig. 5. The main steps of the analogous 'first reaction' algorithm are as follows. A random number $x \sim U(0, 1)$ is drawn and the sojourn time t_n is solved for from $-\ln x = R(t_n)$ (see Fig. 5). The number (M_n) and identities of channels, participating in the next reaction, are defined as all the channels with the activation times satisfying $t_{ij}^a \leq t_n$ ($M_n = 5$ in Fig. 5). The first reaction channel is then determined, using the second draw of a random number, and a discrete M_n-point distribution, as before.

3   Basic predictions and analysis

3.1   Useful metrics

Although bike-sharing is generally designed with short trips in mind, a long bike-sharing trip may be perceived favourably by the user if the alternative would result in an even longer trip. The trip durations, which are the primary data directly measured and modelled, are therefore not adequate to describe the user perspective, unless the effects of competing modes of transport are taken into account. We propose the following three metrics which make such comparison indirectly and are therefore better indicators of a system's functioning than trip durations.

3.1.1   Median trip efficiency

The efficiency, gain, or reward of a trip is a measure of how useful a particular trip is to an agent, as compared with the same trip made by walking. We will use a median efficiency $\mathbb{M}e$, defined through

$$\Pr\{e < \mathbb{M}e\} = \frac{1}{2} \tag{8}$$

to describe the same for a population of agents: bike-sharing trips with larger $\mathbb{M}e$ tend to save travel time with respect to trips with smaller median efficiency.

---

[6]interpreted as rolling an M-faced 'loaded die' with weights $(\hat{\lambda}_i)$.
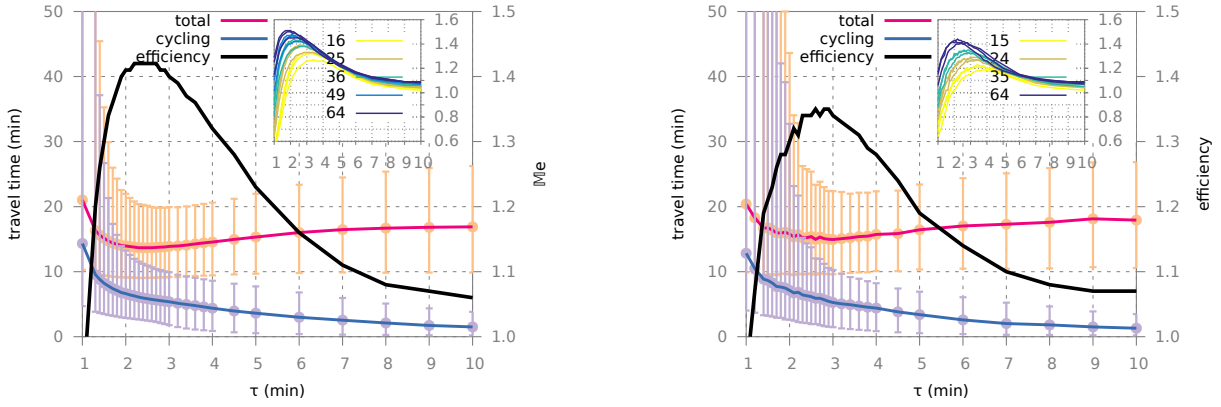
Figure 6: The decision parameter $\tau = 2.5$ yields optimal efficiency, shortest total trips, and the best confidence (central 80% percentile) for the rectangular 6×6 station array of stations (left, cf. Fig. 3). Similar qualitative picture but worse quantitative characteristics (lower efficiency, longer trips, higher uncertainty) characterise the star-shaped 5×7 configuration (right, cf. Fig. 3). Both insets show variations with respect to the number of stations (S) and different filling degrees $N/C = 0.25, 0.5, 0.75$.

3.1.2 Excluded population metric

Agent locations are sampled with the assumption that each agent is interested in using bike-sharing. However, an agent uses bike-sharing only if the triangle inequalities (5), and the cycling tolerance inequality (6), are satisfied. The fraction of all agents who, based on the union of these inequalities do not, or cannot choose bike-sharing (because of empty stations), is called the excluded population metric (EPM).

3.1.3 Congestion metric

In a queueing model approach to bike-sharing, Fricker and Gast [FG14] identify completely empty or full stations as problematic because they inhibit one direction of traffic. To relate their work to ours, we introduce the bicycle and the slot congestions as follows,

$$p_t^+ = \frac{1}{tS} \int_0^t \sum_{i=1}^S 1_{\{s_i(t)=c_i\}}, \quad p_t^- = \frac{1}{tS} \int_0^t \sum_{i=1}^S 1_{\{s_i(t)=0\}}, \tag{9}$$

Since both parameters range from $p_t^\pm = 0$ (no station is ever full/empty), to $p_t^\pm = 1$ (all stations are always full/empty) they give an average measure of 'problematic stations' in the sense of [FG14]. Note that $p_t^+ + p_t^- \leq 1$ for all t.

3.2 Model parameters

The results in this section concern a single agent model (except in section 3.5 where a multi agent model is used) with 6×6 rectangular, or 5×7 star-shaped station configurations as in Fig. 3, with $c_i = 20$ and an initial filling degree $N/C \approx 0.5$, placed in a 3×3 km area (similar to Fig. 11), and random initial configuration of bicycles in stations.

3.3 Decisions determining travel efficiency

The trip efficiency metric (7) provides insight in the effectiveness of bike-sharing trips as shown by the simulation results in Fig. 6. They show that there exists a compromise value of the decision scale parameter τ (see Sect. 2.3.3), which minimises both the median of total trip times and the scatter
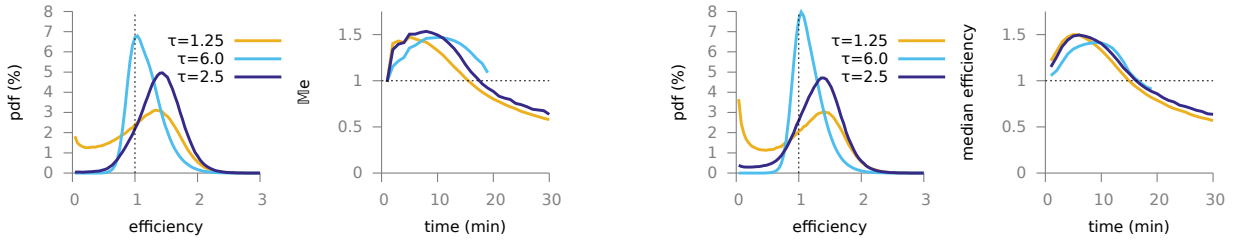
Figure 7: The PDF of trips as a function of the efficiency and median efficiency of trips as a function of cycling time for square shaped configuration (left two graphs, cf. Fig. 3) and a star-shaped configuration (right two graphs, cf. Fig. 3) provide complementary information about the trip–efficiency relation.

of their distribution and maximises the median efficiency. The insert shows the median efficiency for different system sizes and N/C.

Additional insights into the trip–efficiency relationship are provided by Fig. 7 (regular and star-shaped grid examples), showing the PDFs of efficiency, and median efficiency as a function of cycling trip duration for three cases: the near-optimal $\tau = 2.5$ (cf. Fig. 6), and sub-optimal ones: $\tau = 1.25$, and $\tau = 6$. The sub-optimal PDFs either have a peak near $e = 1$, meaning that a typical trip either is not worth taking, or is marginally so ($\tau = 6$), or have excessive exposure to anomalous long trips, accumulating near a zero efficiency ($\tau = 1.25$). Figure 7 (right) shows that travel is most efficient in a window of cycling times between roughly 5 and 15 minutes. The latter time can be explained by the considered size of the area (3×3 km).

Values of τ below a certain threshold may result in the perceived utility of all stations becoming negligible. In that case, the model predicts very long 'cycling' trips, even longer than cycling across the entire city! Curiously, there exists a real life analogy of the negligible utility setting[7]. Occasionally, users abandon their bicycles 'on the curb', preferring to leave them unguarded (paying fees for extended usage, or even a fine) rather than taking time to park them.

3.4 Typical cycling time distributions: algebraic 'tail', major and minor modes

3.4.1 Conditional expectation of travel times

In Fig. 8 the distribution of cycling times is shown for several values of τ. Note that for smaller values of τ, the times are distributed asymptotically as t^{-1-a}, i.e. the distribution has 'an algebraic tail'. Since walking is less uncertain, the distribution of late arrival times to the destination has qualitatively similar asymptotic properties as cycling time distributions. Let us briefly summarise the implications of distributions having algebraic tails, and how the existence of such tails in the cycling distributions may suggest quantifying user dissatisfaction with the system.

Someone interested in the expectation of being late at a destination, would assume that the arrival time is some stochastic process T, with a distribution $F(t)$, i.e. $\Pr\{T < t\} = F(t)$, and would compute the conditional expectation of arrival time, given that currently at time t the destination hasn't been reached yet: $\mathbb{E}_t T = \int_t^\infty y F(dy) / \int_t^\infty F(dy)$. Freely adjusting the reference frame so that the expected arrival time is set to zero, then if $t < 0$ one is early, and if $t > 0$ one is running late. As a general feature of typical distributions, $\mathbb{E}_t T \approx 0$ if $t \ll 0$, meaning that one is expected to arrive 'on time' provided a long enough time allowance. However, if $t > 0$ then the expectation is at least t (meaning $\mathbb{E}_t T \geq t$), but the precise expression depends crucially on F. If the distribution is strictly algebraic 'in the tail', then $\mathbb{E}_t T = t + \frac{1}{a-1}t$ if $a > 1$, and $\mathbb{E}_t T = \infty$ otherwise. Figure 8 (right) shows how $\mathbb{E}_t T$ changes with t for $\tau = 2.5$ leading to on average shorter trips but occasionally long delays, and $\tau = 6$ (more tolerance), leading to on average longer trips, but more predictable delays that are closer to a normal distribution. It is useful to think in terms of the expected delay δ, defined for $t > 0$

---

[7]This was reported to us by the office running the shared bicycle system in Pisa (M. Bertini, private communication)
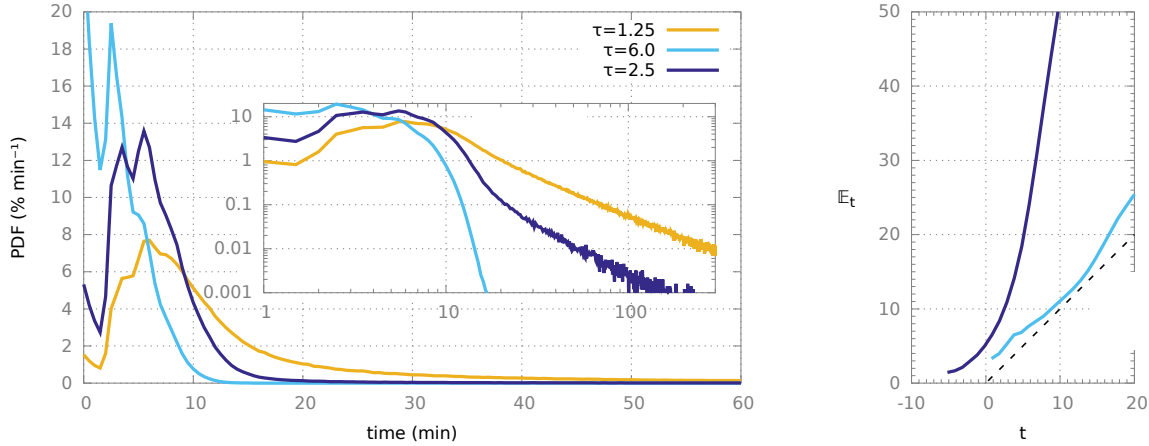
Figure 8:   Cycling time PDF (left) and expected late arrival time (right)

through $\mathbb{E}_t T = t + \delta$. For comparison, the normal distribution yields $\delta \approx \sigma^2/t$ for large t, which means that the near-immediate arrival becomes more certain as the delays accumulate. This is in stark contrast with algebraic distribution, for which we have $\delta = t/(a-1)$. In this case, because $\delta \propto t$, the near-immediate arrival becomes less probable as the delays accumulate. This situation suggests that relatively long undesired trips can sometimes occur that would be perceived negatively by the users who, as we stipulated in the beginning, are time-conscious agents. The exponent a, or the ratio $a/(a-1)$ can be used to quantify this effect. A qualitative comparison of the curves in Fig. 8 (left) and the histograms in Fig. 1 (Data) suggest that the actual behaviour is consistent with moderate risk-taking, exemplified by the optimal $\tau = 2.5$.

3.4.2   Rational vs. smart decisions

The model predicts a minor mode at zero minutes' travel. This feature is present in some cities (Pisa, see below, and Lyon, see [BAF+11]) but is absent in the London dataset. Borgnat et al. [BAF+11] suggest that mistakenly hired malfunctioning vehicles are the explanation. In our model, short trips result invariably from a station of a previous hiring being selected for returning. Such trips exist in the model because all stations, including ones that have just been used for hiring, are equally valid (but not equally probable) options for returning. This is a consequence of the distance-only dependent utility model. A typical setting, whose likely outcome is a 0-minute trip, should have a common nearest station to both the origin and destination locations, placed roughly in the middle between the two objectives. In this setting, it would be 'rational' to hire a bike at the midway station, only to find afterwards that it is also rational to return it there too, even if an agent would be better off walking to the destination directly. This feature of the model is kept because it reproduces the data in Pisa while the overestimate in the London dataset is moderate, and because we have no factual data to rule out a possibility of similar decisions by real users. In the case of a future confirmation of a plausible hypothesis by Borgnat et al., it could be changed by a modification to the station utility model.

3.4.3   Central mode of the distribution

The most prominent feature of the distributions is a mode, typically around 10 minutes. We found that its precise location and the shape of the distribution in the vicinity of the mode is approximated to a good degree by assuming that the cycling tolerance parameter t_c (see section 2.3.4) is a random number from a uniform distribution, $t_c \sim U(t_{\min}, t_{\max})$. Presence of the cutoff in the expected cycling time is an important underlying property leading to the observed distributions, affecting the characteristic 'neck' of these distributions between the mode and the crossover into the algebraic 'tails' Fig. 1.
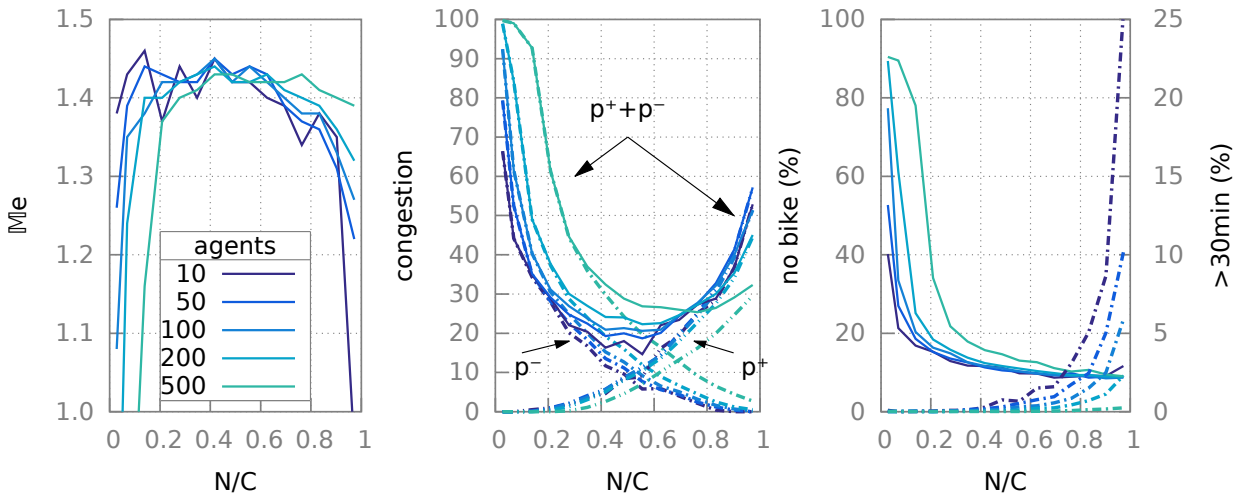
Figure 9: System characteristics as a function of the stations filling degree N/C.

Typically in real systems, $t_{\max}$ corresponds to the free cycling allowance.

3.5   Travel efficiency describing the agent's perspective

In a queueing model approach to bike-sharing, Fricker and Gast [FG14] predict that the population of problematic stations increases when a system has many, or few bicycles with respect to its capacity. They show the existence of an optimal station filling degree, whose value is not far from $N \approx \frac{1}{2}C$. In our model, neither of the used metrics, except for $p^+ + p^-$ demonstrate conclusively relevance of such an optimality criterion. For example, the efficiency (and with it user satisfaction) remain high in a wide range of occupancy numbers without a statistically significant extremum. The median efficiency, taken as a function of the filling level is a highly fluctuating variable, as suggested by Fig. 9 (left)[8]. As another example, Fig. 10 shows that the excluded population remains low for a larger value of τ. The latter would suggest an equally high service level by a network of any size, provided that agents are willing to search for a station in a large area. However, as we have seen in Fig. 6, the efficiency deteriorates markedly with increasing τ. The apparent disagreement with results in [FG14] is mostly in the interpretation of the term 'problematic'. In our model, no normative assumptions regarding the 'fitness' of the stations separately from the agent dynamics are being made. Instead, performance of a system is evaluated on the trips that actually take place and by gauging their effectiveness against certain benchmark durations available for each trip. In this view it is then not surprising that, for example, agents, who manage to use bike-sharing in a nearly empty system, need not find their experience 'problematic'. On the other hand, an empty system leads to high EPM (Fig. 9, right) which should be troublesome news to an operator, committed to increasing a system's usage. This and the large fraction of trips in the 'tail' of the distribution when a system is too full (Fig. 9, right) should be alarming to potential users who do not find a suitable bike to hire or return. Increasing the number of agents will increase the flows, and deteriorate the statistics of user-side efficiency. There may be conflicting objectives addressing different 'problematic' aspects. We believe that similar conflicting values maybe among the issues, faced or to be faced by designers of real systems, and it is therefore desirable that an adequate model is not limited to predicting a unique optimality criterion. Exploring several metrics with our models we may begin to address issues like

- Given a certain density of population, presumably interested in using the bike-sharing, and a certain configuration of bicycle stations, what is the fraction of population that will find it impractical to use it?

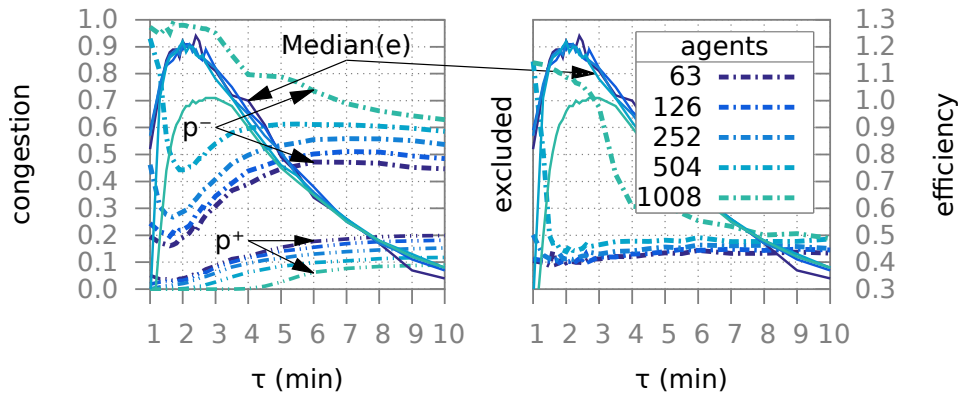[8]near empty and near full system extremes are clearly unfavourable

Figure 10: Model with $N = 252$ and $C = 720$: Median efficiency is relatively insensitive to rising congestion levels (left panel). Likewise the EPM, although rigid tolerance (small τ) result in an unusable system (right panel).

- Can efficiency of trips be improved by changing the number of vehicles or station capacities?

Our model shows that these objectives are not identical and possibly require a trade-off. However, they suggest several ways of attacking the problem such as considering different system configurations or changing the utility perception by introducing incentives. We have shown that such models can address these issues, considering both potential and actual users of the system, providing insight into system performance from a user point of view that is complementary to pure data analysis of real systems. The latter naturally excludes data on potential but unsuccessful users which is hard to obtain in other ways.

4 Pisa and London datasets

The bike-sharing systems in London and Pisa differ in size by orders of magnitude, as is evident from Table 1.

Characteristic	Pisa	London
Stations	15	742
Approximate fleet size	150	11,500
Approximate capacity	270	19,000
Average trips per hour	32	1120
Approximate area (km^2)	15	90
Average stations (km^{-2})	1	8
Cycling $\geq$ 30 min (%)	6.0	7.7

Table 1: London and Pisa bike-sharing systems

And yet, the cycling time distributions in these cities bear distinctive similarities, as shown by the filled areas in Fig. 1, shared also by other cities [BAF+11]. It is tempting to consider them as members of a family of distributions, characterised by a major mode at about 10 minutes, algebraic tails of the distributions with exponent $a \approx 2$, containing about 7% of trips longer than 30 minutes, and a minor mode close to 0 minutes (the latter is absent in the London data).
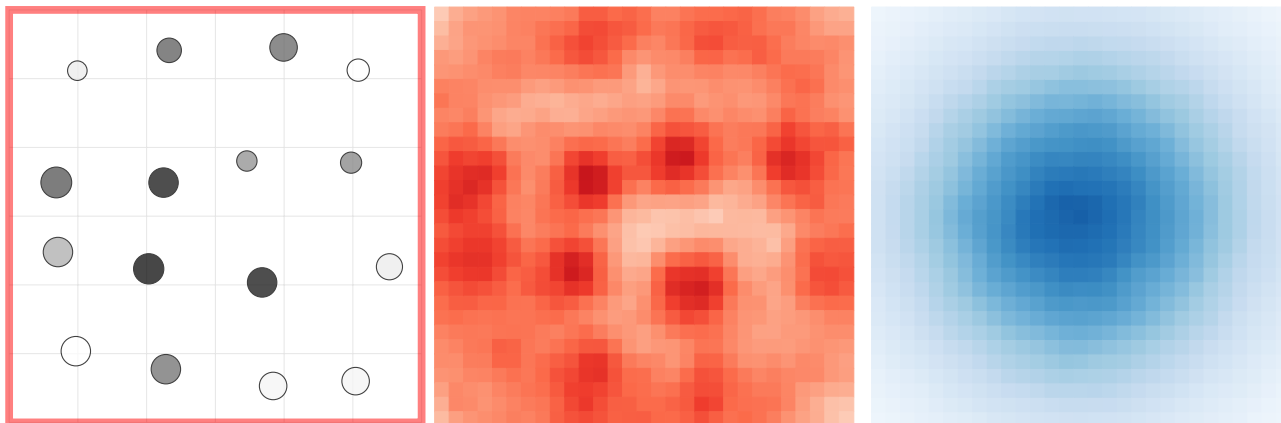
Figure 11: Spatial set-up to simulate the Pisa data-set. Left panel: The map of stations and a random snapshot of their filling (circle size $\propto c$, shade $\propto n$). Middle, right: distributions of the origin and destination locations, respectively.

4.1   Model setup

It was suggested that at least three kinds of agents would be required to model similar distributions [BAF+11]. In the following we show that a single type of agent is sufficient to generate qualitatively correct distributions with all aforementioned characteristics. This can be achieved without sophisticated parameter fitting, by using only some qualitative arguments. We emphasise that the objective of this study is not a model that copies a real system in question; rather, it is geared towards a model that incorporates quantitatively correct features of user behaviour, with the aim of providing insights into the plausible underlying reasons for the observed data. In fact, the simple models discussed section 3 already contain many of the salient features, although they are not sufficiently accurate for what concerns some quantitative aspects.

The curves for the uniform models, shown in Fig. 1, correspond to a multi-agent model (see Table 2) with uniformly distributed agent locations. Such distributions generate statistical self-redistribution of bicycles. As such, it is a statistically optimal system.

The single necessary additional feature to obtain good quantitative agreement between collected data and model results such as those in Fig. 1, are the agent flows. Their introduction lead to some areas being consistently short of bicycles, and others runing out of available parking slots. Such flows can be easily constructed using inhomogeneous spatial distributions, and they may be given an additional temporal dimension, for example by requiring agents to honour synchronised appointments. Such temporal 'tidal flows' are present in real systems in the form of morning and afternoon commutes in opposing directions, often with a clear spatial separation. It was found by experimentation that temporal features accentuate the effects of scarce resources and that this effect is quantitatively similar to the static flows.

The model of Pisa represents a relatively small system with a 4×4 array of stations (see Table 2). Calibration with a single-agent model, as in Fig. 6, yields $\tau^o \approx 3.5$ (min), which is quite large as a result of sparsely distributed stations. To introduce flows, it is sufficient to sample the origin locations from a uniform distribution, and the destinations from a Gaussian distribution, concentrated in the centre[9]. In this way, the periphery is tendentially lacking bicycles, most of which have been moved towards the centre (see Fig. 11).

The model of London covers a larger area with a 19×38 array of stations (see Table 2). Calibration as above yields $\tau^o \approx 1.8$ reflecting a denser network. Introduction of multiple centres of hiring and returning, approximately the size as in Pisa and of counter-current flows for the balancing, achieves

---

[9]The discrepancy between these trial and actual distributions visible in Fig. 11 are the result of rejected trips, see sections 2.3.4 and 3.1.2

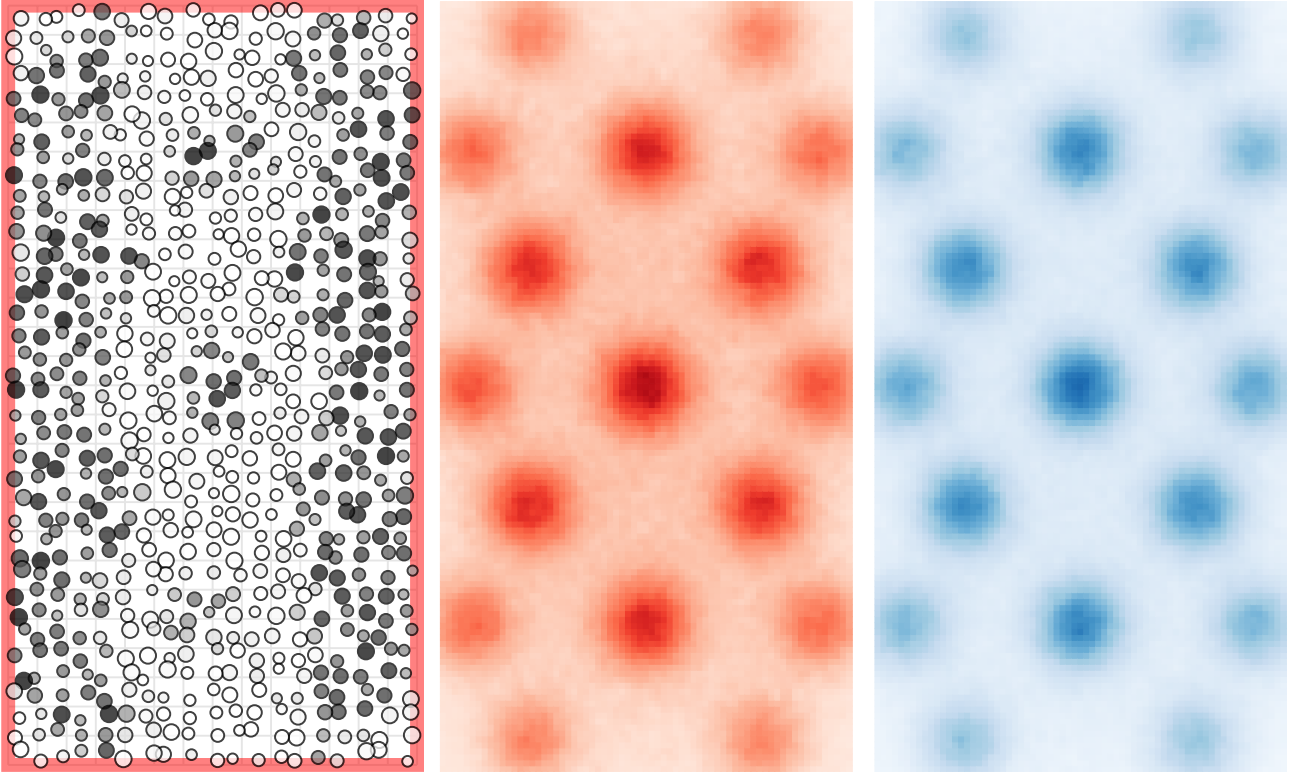| Model	τ^o (min)	Area + stations (km²)	S	Capacity min/max	C	Bikes N	Trips (h^{-1})
Pisa	3.5	3×3	16	10/25	170	304	35
London	1.8	7×13	722	15/40	19,652	10,000	823

Table 2: Model parameters

Figure 12: Spatial set-up to simulate the London data-set. Left panel: The map of stations and a random snapshot of their filling (circle size $\propto c$, shade $\propto n$). Middle, right: distributions of the origin and destination locations, respectively.

the result shown in Fig. 12.

The simulation results for both models are presented in Fig. 1 (the 'flow model'). Note that there are trips lasting longer than 30 minutes. Since agents use bike sharing with an *a-priori* expectation that their trip should be shorter than a certain $t_{\max}$, with sup $t_{\max}$ being the aforementioned figure, we know that all trips 'in the tail' were not intended to be so long. This is a first qualitative indication of possible dissatisfaction on the user side. Comparing the uniform vs. flow models: The number of trips in which agents do not find a bike is 10% (Pisa) and about 1% (London). Trips in the tail make up 2 vs. 5 % (Pisa) and 2 vs. 7.7% (London). The completely full stations are 12 vs. 25% (Pisa) and 10 vs. 20% (London). Clearly, longer trips are positively correlated with full stations. This suggests that full stations are indeed a plausible cause of the occurrence of most likely undesired longer trips, and thus a source of concern for user satisfaction. Such concerns have also been expressed in an on-line survey conducted in 2009 among users of the bicycle sharing system in Barcelona [FNO09]. Of the 212 respondents 76% mentioned *finding an available bicycle* as an important problem in their experience and 66% mentioned *finding an available parking slot*.

Based on the estimated value of the exponent a for both data-sets ($a \approx 2.1$ for Pisa and 1.9 for London), the performance of the two systems in this respect is similar. However, comparison between
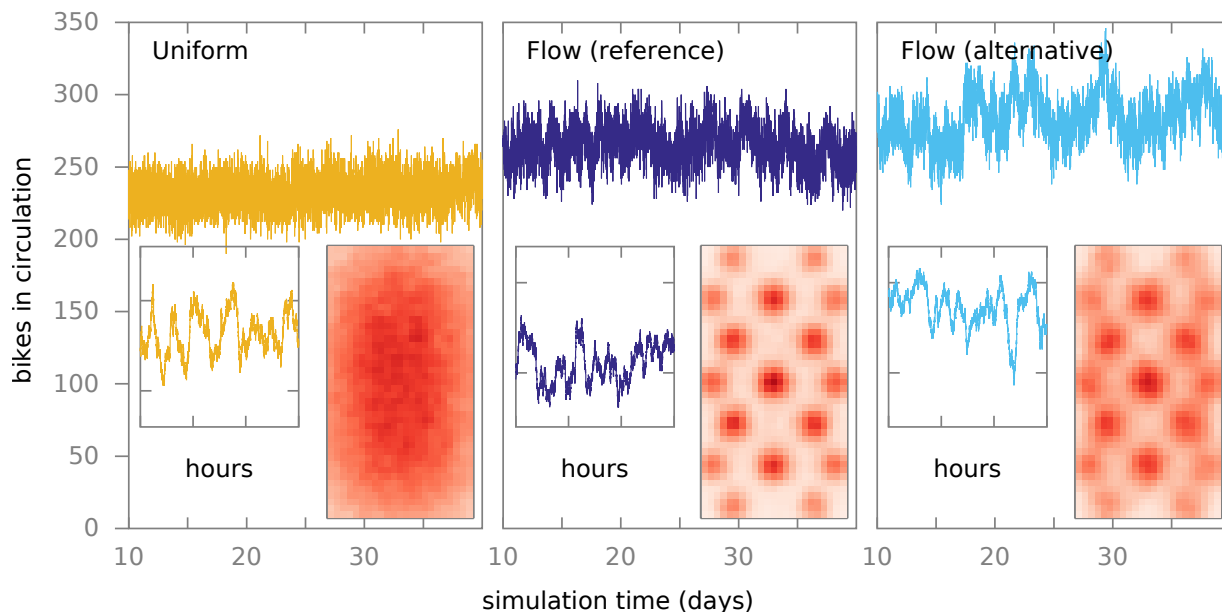
Figure 13: Bicycles in circulation for the uniform flow model (left), the reference flow model (middle) and the alternative flow model (right).

the data and the corresponding uniform model (optimal in a sense described before) shows only a small difference between the two distributions for Pisa, and a bigger difference for London (a 5 times smaller PDF in the tail compared to the data). This suggests that the performance in Pisa cannot be improved much unless more stations are added. However, in London, which has an 8 times higher density of stations, the situation is not significantly better. The model suggests that there are areas with predominantly full stations (somewhat like those in Fig. 12) rather than isolated full stations. This is most likely the main explanation for the number of unintentionally long trips.

4.2 Time scale(s) of bicycle circulation

The aggregate number of bicycles in circulation in a system is a fluctuating quantity. There two types of fluctuations: stochastic and tidal.

4.2.1 Stochastic fluctuations

Even when the distribution of agent objectives is stationary, the number of bicycles in circulation fluctuates because the agent behaviour is probabilistic. The parameters of fluctuation amplitude or the variance are to be reckoned with: an operator may get more repositioning requests if the variance is larger or if the fluctuations lead to long-time trends, or large and unexpected excursions from the mean. Simulations, extending to very long times are useful in that respect, since they give an idea about the statistics of fluctuations.

Simulations of the London setup demonstrate that there is a significant dependence of fluctuations on the spatial distributions of agents. Figure 13 shows the number of bicycles in circulation for the London model that is parameterised by different spatial distributions of agents returning bikes. The uniform model, the reference flow model, and the alternative flow model generate quite different fluctuation magnitudes, discernible in Fig. 13. These fluctuations appear qualitatively similar on a three-hour scale (insets) but are strikingly different on a ten-day scale: from stationary-appearing fluctuations in the uniform model, to non-stationary for the reference flow model used in Fig. 1, to even
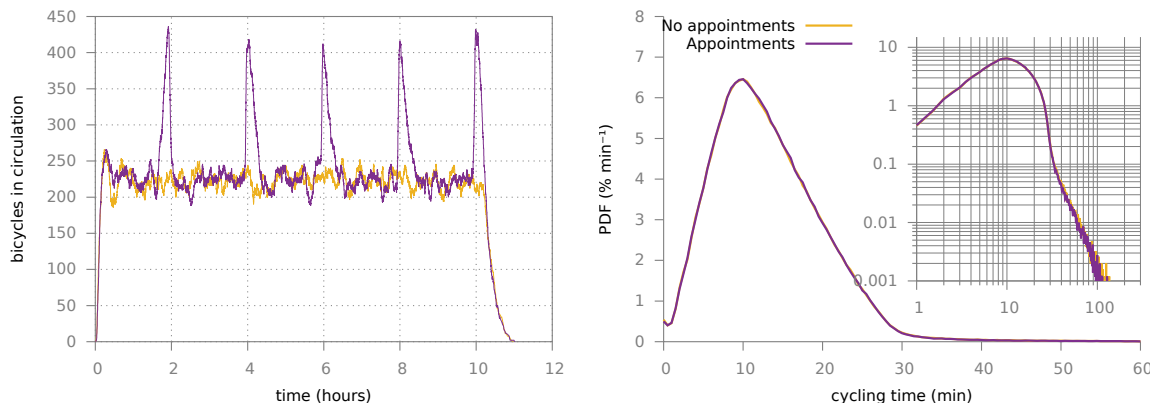
Figure 14: Uniform London model with and without cyclic synchronous appointments.

more volatile case of the alternative flow model (Fig. 13: left, middle, and right panels, respectively).[10] Similarities on a three-hourly scale and differences on a ten-day scales suggest that spontaneous vehicle fluctuations contain a significant multi scale component. Moreover, even the three-hourly insets show some evidence of sub-hourly trends which is surprising given that most trips are less than 30 minutes long. Note that the term 'multi scale' here does not refer to the organisational multiple scales (such as various temporal horizons of planning activities) but to 'fluctuational' multiple scales. Examples of systems with multiple time scales abound in the literature, showing that even a relatively simple model structure can yield a complicated structure of fluctuations such as, for example, intermittency. Generically, the manifestations of multi scale behaviour should be considered as contingencies of additional higher 'risk'. In the case of bike-sharing model these risks are embodied by areas with limited vehicle availability which are larger than what could be expected from the uniform model.

4.2.2   Cyclic fluctuations

Because people in real cities are engaged in activities with similar schedule times, such as the beginning, or the ending of a working day or studies, their activies are synchronous to some extent. Such behaviour causes spikes in the traffic flows and can be thought of as the result of certain cyclic 'tidal flows' in the form of morning or afternoon commutes in opposing directions, often with a clear spatial separation. They induce large and quite predictable temporal variation in the circulation, described as cyclostationarity [GNP06]. Cyclic flows can be included in the model. For example, Fig. 14 shows the number of bicycles in circulation on a smaller scale for the uniform London model. In the first model all agents have randomly distributed appointments (as previously). The second model has two groups of agents. The first group has random appointments, whereas the second group must honour appointments, set at two hour intervals apart. Relatively large populations rushing to their targets at the same time causes visible peaks in Fig. 14 (left), whereas the cycling time distribution for both groups, shown in Fig. 14 (right), are essentially the same.

5   Conclusion

The inclusion of minimal, but plausible, user behaviour and flows in a general bike-sharing model based on MRPs is shown to be sufficient to explain some of the main features of the distributions of actually observed data. The approach provides complementary insight into the attractiveness of bike-sharing from a user's perspective, including that of potential users not captured by data sets. Future work will address its use for the evaluation of the effects of alternative configurations and

---

[10]The last setup is similar to the reference flow model, except that the areas of arrival requests have been enlarged

user incentives. Furthermore, we plan to compare our results with approximate mean-field models of bike-sharing [LLM15].

Acknowledgement

References

[And07]   D. F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J. Chem. Phys.*, 127(21):214107, 2007.

[BAF+11]   P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury. Shared bicycles in a city: A signal processing and data analysis perspective. *Adv. Complex Syst.*, 14(3):415–438, 2011.

[BMLEG12] J. Bachand-Marleau, B. H. Y. Lee, and A. M. El-Geneidy. Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. *Transport. Res. Rec.: J. TRB.*, 2314(1):66–71, 2012.

[Car14]   A. Carey. Spoke too soon: Melbourne bike share to drag chain another year. *The Age*, Sept 4, 2014.

[Ç75]   E. Çinlar. *Introduction to stochastic processes*. Prentice-Hall, 1975.

[Chh88]   R. Chhikara. *The Inverse Gaussian distribution: Theory, Methodology, and Applications*, volume 95. CRC Press, 1988.

[FG14]   C. Fricker and N. Gast. Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *EURO J. Transport. Log.*, pages 1–31, 2014.

[FNO09]   J. Froehlich, J. Neumann, and N. Oliver. Sensing and predicting the pulse of the city through shared bicycling. In *IJCAI*, pages 1420–1426, 2009.

[FWH14]   E. Fishman, S. Washington, and N. L. Haworth. Bike share's impact on car use: evidence from the United States, Great Britain, and Australia. In *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, 2014.

[GB00]   M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104(9):1876–1889, 2000.

[GHP13]   D. T. Gillespie, A. Hellander, and L. R. Petzold. Perspective: Stochastic algorithms for chemical kinetics. *J. Chem. Phys.*, 138(17):170901, 2013.

[Gil76]   D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22(4):403–434, 1976.

[GNP06]   W. A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: Half a century of research. *Signal processing*, 86(4):639–697, 2006.

[Kah03]   D. Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *Am. Econ. Rev.*, pages 1449–1475, 2003.

[Kul95]   V. G. Kulkarni. *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, 1995.

[LLM15]   D. Latella, M. Loreti, and M. Massink. On-the-fly PCTL fast mean-field approximated model-checking for self-organising coordination. *Science in Computer Programming*, 2015. Accepted for publication.

[Mai09]   P. De Maio. Bike-sharing: Its history, impacts, models of provision, and future. *J. of Public Transport.*, 12(4):41–56, 2009.

[Mid11]   P. Midgley. Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas. In *19th session of the Commission on Sustainable Development*, CSD19/2011/BP8. United Nations, 2011.

[MP09]   J. H. Miller and S. E. Page. *Complex adaptive systems: an introduction to computational models of social life*. Princeton, 2009.

[OCB14]   O. O'Brien, J. Cheshire, and M. Batty. Mining bicycle sharing data for generating insights into sustainable transport systems. *J. Transport Geogr.*, 34:262–273, 2014.

[PDH10]   J. Pucher, J. Dill, and S. Handy. Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive medicine*, 50:S106–S125, 2010.

[tBFG14]   M. H. ter Beek, A. Fantechi, and S. Gnesi. Challenges in modelling and analyzing quantitative aspects of bike-sharing systems. In *Leveraging Applications of Formal Methods, Verification and Validation*, volume 8802 of *LNCS*, pages 351–367. Springer, 2014.