User-based solutions for increasing level of service in bike-sharing transportation systems

Juste Raimbault[1,2]
[1]Erasmus Mundus Masters in Complex Systems Science, Ecole Polytechnique ParisTech
[2]LVMT, Ecole Nationale des Ponts et Chaussées

December 15, 2013

⋆ ⋆
⋆

Erasmus Mundus Masters in Complex Systems Science
Open Problems Project

Supervisor: Khashayar Pakdaman, Institut Jacques Monod, CNRS UMR 7592, Paris
Advisor: Arnaud Banos, ISCPIF and CNRS UMR Géographie-Cités, Paris

⋆ ⋆
⋆

Abstract

City Bikes sharing transportation systems have been well studied from a top-down viewpoint, either for an optimal conception of the system, or for a better statistical understanding of their working mechanisms in the aim of the optimisation of the management strategy. Yet bottom-up approaches that could include behavior of user have not been studied well so far. We propose an agent-based model for the short time evolution of a bike-sharing system, with a focus on two strategical parameters that are the role of the quantity of information users have on the all system and the propensity of user to walk after having dropped their bike. We implement the model in a general way so it is appliable to every system as soon as data are available to a certain standard. The model of simulation is parametrized and calibrated on processed real time-series. After showing the robustness of the simulations regarding random parameters, we are able to test different user-based strategies for an increase of the level of service. In particular, an increase of user information can have significant impact on the homogeneity of repartition of bikes in station. An increase of propensity to walk can also have positive effect, but that stays a parameter that is difficult to change in reality.

1 Introduction

Sharing bikes transportation systems, or also called "smart bikes", have been presented as an ecological and user-friendly transportation mode, which appears to be well complementary to classic public transportation systems ([1]). The quick propagation of many implementations of such systems across the world can be seen as confirmation of the interesting potentialities that bike can offer, as it is synthetised in [2]. O'Brien & *al.* propose in [3] a global vision on the current state of bike sharing across the world.

Inspired by the relatively good success of such systems in Europe, the possible key factors for their quality have been questionned and tranposed to different possible countries of implementation such as China ([4, 5]) or the United States ([6]).

The understanding of how a system works is essential and can be done through statistical analysis with elaborated statistical models ([7, 8, 9, 10]) or data-mining techniques ([3, 11]), and can lead to the understanding of broader questions such as structure of mobility patterns.

Concerning the implementation, a crucial point in the design of the system is an optimal location of stations. That problem have been extensively studied from an Operationnal Research point of view ([12, 13] for example). The next step is a good exploitation of the system. By nature, strong dissymetries appear in the distribution of bikes, translating in a way urban mobility patterns. That becomes in many cases problematic because leads to a string decrease in the level of service (no parking places or no available bikes for example). To counter such phenomena, operators have redistribution strategies, that have also been studied a lot and for which opimal redistribution strategies have been proposed ([14, 15, 16]).

However, all these studies always approach the problem from a top-down point of view, and bottom-up strategies have been to our knowledge poorly considered in the litterature. User-based methods have been considered in [17, 18] in the case of a car-sharing system, but the problem stays quite far from a behavioral model of the agents using the system, since it explores the possibility of implication of users in the redistribution process, or of shared travels what is not relevant in the case of bikes. Indeed the question of a precise determination of users behaviors and parameters on the level of service of a bike-sharing systems remains open.

Therefore we propose an agent-based model of simulation in order to represent and simulate the system from a top-down approach, considering bikers and parking as stations as agents and representing their interactions and evolutions in time.

2 Presentation of the model

Introduction The granularity of the agent-based model is at the scale of the individual biker and of the stations where bikes are parked. A more integrated view such as flows would not be useful to our purpose since we want to study the impact of the behavior of individuals on the overall performance of the system. The global working scheme consists in agents embedded in the street infrastructure, interacting with particular elements, what is inspired from the core structure of the Miro model ([19]). The scale of the space is globally one of a district; we don't consider the whole system for calculation power purposes (around 1300 stations on all the system of Paris, whereas an interesting district have around 100 stations), what should not be a problem as soon as in- and outflows are well considered. It is also a deliberate choice because we want to focus on some districts where repartition problems are particularly important and not risking to loose information about these critical cases by taking the system in its whole. Time scale of a run is logically one full day because of the cyclic nature of the process ([20]).

Formalisation The street network of the area is an euclidien network ($V \subset \mathbb{R}^2, E \subset V \times V$) in a closed bounded part of $\mathbb{R}^2$. The time is discretised on a day, so all temporal evolution are defined on $T = [0, 24] \cap \tau\mathbb{N}$ with τ time step (in hours). Stations S are particularly vertices of the network for which constant capacities $c(s \in S)$ are defined, and that can contain a variable number of bikes $p_b(s) \in \{0, \ldots, c\}^T$. We suppose that temporal fields $O(x, y, t)$ and $D(x, y, t)$ are defined, corresponding respectively to probabilities that a given point at a given time becomes the expected departure (resp. the expected arrival) of a new bike trip, knowing that the trip exists. Boundaries conditions are represented as a set of random variables $(N_I(i, t))$ that for each possible entry point $i \in I$ ($I \subset V$ is a given set of boundaries points) and each time, $N_I(i, t)$ gives the number of bikes trips entering the zone at point i and time t. For departures, a random time-serie $N_D(t)$ represents the number of departures in the zone at time t. Note that these random variables and probabilities fields are sufficient to built the complete process of travel initiating at each time step. Parametrisation of the model will consist in proposing a consistent way to construct them from real data.

The stations are fixed agents, only their functions p_b will vary through time. The other core agents are the bikers, for which the set $B(t)$ is variable. A biker $b \in B(t)$ is represented by

- its mean speed $\bar{v}(b)$

- a distance $r(b)$ corresponding to its "propensity to walk". This is a key parameter that represents the distance he is ready to walk between the station where we will take/leave his bike and his actual objective.

- a boolean $i(b)$ expressing the capacity of having access to information on the whole system at any time (through a mobile device and the dedicated application for example)

We are then able to define the workflow of the model for one time step. The following scheme is sequentially executed for each $t \in T$, representing the evolution of the system on a day.

For each time step the evolution of the system is done by:

- Starting new travels. For a travel within the area, if biker has information, he will adapt his destination to the closest station of its destination with free parking places, if not his destination is not changed.
  - For each entry point, draw number of new traveler, associate to each a destination according to D and characteristics (information drawn uniformally from proportion of information, speed according to fixed mean speed, radius also).
  - Draw new departures within the area according to O, associate either destination within (in proportion to a fixed parameter p_{it}, proportion of internal travels) the area, or a boundary point (travel out of the area). If the departure station is empty, biker walks to an other station (with bikes if has information, a random if not) and will start his travel after a time determined by mean walking speed and distance of the station.
  - Make bikers waiting for start for which it is time begin their journey (correspond to walkers for which a departure station was empty at a given time step before)

- Making bikers for who travel is in progress advance of the distance corresponding to their speed

- Finishing travels or redirecting bikers
  - if biker was doing an out travel and is on a boundary point, travel is finished (get out of the area)
  - if has no information and is not on a station, go to a random station within $r(b)$
  - if is on a station with free places, drop the bike
  - if is on a station with no places, choose as new destination either the closest station with free places if he has information, or a random one within $r(b)$ (excluding already visited ones, implying the memory of agents).

Fig. 1 shows the decision process for starting and arriving bikers. Note that walking radius $r(b)$ and information $i(b)$ have implicitly great influence on the output of the model, since droping station is totally determined (through a random process) by these two parameters when the destination is given.

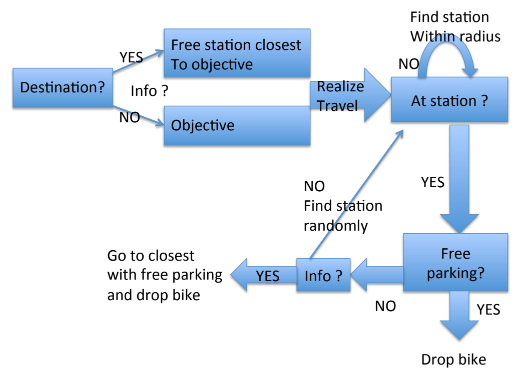


Fig. 1 | Decision process of bikers.

Evaluation criteria In order to quantify the performance of the system, to compare different realisations for different points in the parameter space or to evaluate the fitness of a realisation towards real data, we need to define some functions of evaluation, proxies of what are considered as "qualities" of the system.

Temporal evaluation functions These are criteria evaluated at each time step and for which the output on the all shape of the time-series will be compared.

- Mean load factor

$$\bar{l}(t) = \frac{1}{|S|} \sum_{s \in S} \frac{p_b(s)}{c(s)}$$

- Heterogeneity of bike distribution: we aggregate spatial heterogeneity of load factors on each station through a standard normalized heterogeneity indicator, defined by

$$h(t) = \frac{2}{\sum_{s \neq s' \in S} \frac{1}{d(s,s')}} \cdot \sum_{\substack{s, s' \in S \\ s \neq s'}} \frac{\left| \frac{p_b(s,t)}{c(s)} - \frac{p_b(s',t)}{c(s')} \right|}{d(s,s')}$$

Aggregated evaluation functions These are criteria aggregated on a all day quantifying the level of service integrated on all travels. We note $\mathcal{T}$ the set of travels for a realisation of the system and $\mathcal{A}$ the set of travel for which an "adverse event" occured, i. e. for which a potential dropping station was full or a starting station was empty. For any travel $v \in \mathcal{T}$, we denote by $d_{th}(v)$ the theoretical distance (defined by the network distance between origin and initial destination) and $d_r(v)$ the effective realised distance.

- Proportion of adverse events: proportion of users for which the quality of service was doubtful.

$$A = \frac{|\mathcal{A}|}{|\mathcal{T}|}$$

- Total quantity of detours: quantification of the deviation regarding an ideal service

$$D_{tot} = \frac{1}{|\mathcal{T}|} \cdot \sum_{v \in \mathcal{T}} \frac{d_r(v)}{d_{th}(v)}$$

- Detours for adverse travels: same as before but integrated only on adverse events

$$D_A = \frac{1}{|\mathcal{A}|} \cdot \sum_{v \in \mathcal{A}} \frac{d_r(v)}{d_{th}(v)}$$

We also define a fitness function used for calibration of the model on real data. If we note $(lf(s,t))_{s \in S, t \in T}$ the real time-series extracted for a standard day by a statistical analysis on real data, we calibrate on the mean-square error on all time-series, defined for a realisation of the model by

$$MSE = \frac{1}{|S|\,|T|} \sum_{t \in T} \sum_{s \in S} \left(\frac{p_b(s,t)}{c(s)} - lf(s,t)\right)^2$$

3 Results

3.1 Implementation and parametrisation

Implementation The model was implemented in NetLogo ([21]) including GIS data through the GIS extension. Preliminary treatment of GIS data was done with QGIS ([22]). Statistical pre-treatment of real temporal data was done in R ([23]), using the NL-R extension ([24]) to import directly the data. See supplementary material S2 (source code) for more details on the implementation.

Concerning the choice of the level of representation in the graphical interface, we followed Banos in [25] when he argues that such exploratory models can really be exploited only if a feedback through the interface is possible. It is necessary to find a good compromise for the quantity of information displayed in the graphical interface. In our case, we represent a map of the district, on which link width is proportional to current flows, stations display their load-factor by a color code (color gradient from green, $lf(s) = 0$, to red, $lf(s) = 1$). Bikes are also represented in real time, what is interesting thanks to an option that allow to follow some individuals and visualize their decision process through arrows representing original destination, provenance and new destination (should be implemented in further work). This feature could be seen as superficial at this state of the work but it appears as essential regarding possible further developments of the project (see discussion section). Fig. 2 shows an example of the graphical interface of the implementation of the model of simulation.

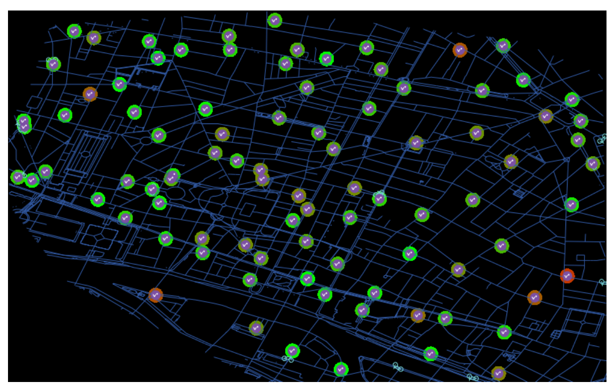


Fig. 2 | Graphical interface of the model for a particular district (Chatelet).

Data collection All used data are open data, in order to have good reproducibility of the work. Road network vector layer was extracted from OpenStreetMap ([26]). Time-series of real stations statuts for Paris were collected automatically[1] all 5 minutes during 3 month (see supplementary material S1 for more details) and were imported into R for treatment with [27] and the point dataset of stations was created from the geographical coordinates with [28].

[1]from the dedicated website api.jcdecaux.com

Parametrisation The model was designed in order to have real proxies for most of parameters. Mean travel speed is taken as 14km.h^{-1} from [29], where data of trips where studied for the bike system of the city of Lyon, France. To simplify, we take same speed for all bikers, a possible extension with tiny gaussian distribution around the mean showed in experiments to bring nothing more. It has been shown in [3] that profiles of use of bike systems stays approximatively the same for european cities (but can be significantly different for cities as Rio or Taipei), what justify the use of these infered data in our case. We also use the determined mean length of travel from [16] (here that parameter should be more sensible to the topology so we prefer extract it from this second paper although it seems to have subsequent methodological bias compared to the first rigourous work on the system of Lyon), which is 2.3km, in order to determine the diameter of the area on which our approach stays consistent. Indeed the model is built in order to have emphasis on travels coming from the outside and on travels going out, internal travels have to stay a small proportion of all travels. In our case, a district of diameter 2km gives a proportion of internal travels of around 20%. We will take districts of this size with this fixed proportion in the following.

The delicate part of the parametrisation is the construction of O, D fields and random variables N_I, N_D from the real data. See supplementary material S1 for a precise mathematical description of used procedures. Daily data were reduced through sampling of time-series of load-factors of all stations and dimension of the representation of a day was significantly reduced through a k-means clustering procedures (classically used in time-series clustering as it is described in [30]). These reduced points were then clustered again in order to isolate typical weekdays from week-ends, where the use profiles are tipically different and from special days such as ones with very bad climate or public transportation strikes. That allowed to create the profile of a “standard day” that was used to infer O, D fields through a Gaussian multi-kernel estimation (see [31]). The characteristic size of kernels $1/\sigma$ is an essential parameter for which we have no direct proxy, and that will be fixed through a simplified calibration. The laws for N_I, N_D were taken as binomial: for an actual arrival, we consider each possible travel and increase the number of drawing of each binomial law of entries by 1 at the time corresponding to mean travel time (depending on the travel distance) before arrival time. Probabilities of binomial laws are $1/Card(I)$ since we assume independance of travels. For departure, we just increase by one drawingsof the binomial law at current time for an actual departure.

What we call parameter space in the following consists in the 3 dimensional space of parameters that have not been fixed by this parametrisation, i. e. the walking radius r (taken as constant on all bikers, as for the speed), the information proportion p_{it} and the size of the Gaussian kernels σ (note that the effective size is $1/\sigma$ so the spread of distribution is decreasing with σ, but we will denote the size that way for convenience in the following).

3.2 Robustness assessment, exploration and calibration

Internal consistence of the model Before using simulations of the model to explore possibe strategies, we need to assess that the results produced are internally consistent, i. e. that the randomness introduced in the parametrisation and in the internal rules do not lead to divergences in results. Simulations were launched on a large number of repetitions for different points in the parameter space and statistical distribution of aggregated outputs were plotted. Fig. 3 shows example of these results. The relative good gaussian fits and the small deviation of distributions confirm the internal consistence of the model.

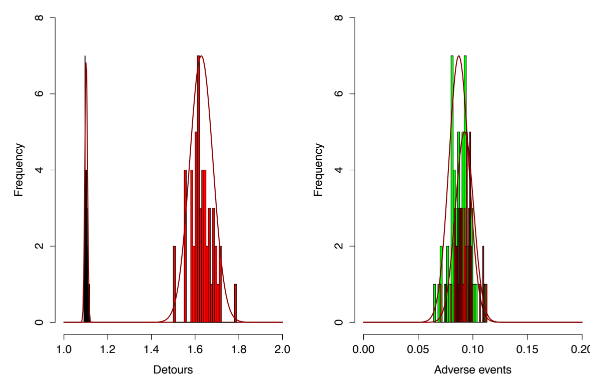


Fig. 3 | Statistical analysis of outputs. For some aggregated outputs (here the overall quantity of detours and the proportion of adverse events), we plotted histograms of the statistical distribution of the functions on many realisations of the model for a point in the paremeter space. Two points are plotted here. The Gaussian fits are also drawn. The relative good fit shows the internal consistence of the model and we are able to quantify the typical number of repetitions needed when using the model: because of the Central Limit Theorem, 100 repetitions are needed to be within $\sigma/10$ for example, but here the relative small values of σ (1/10 of indicator range) suggests quite less to have satisfying results.

Robustness regarding the study area We also need to test the sensitivity of the model regarding geometry of the area. If the structure of the road network and the spatial distribution appear to have a more significant influence on the output than the parametrisation of the model, it would have no sense to evaluate and compare strategies on parameters, since the obtained result would depend on the area on which calculation is done.

Therefore, we compare results obtained for temporal indicators and aggregated ones, on different geometries with the same parametrisation (supposing the same number of station), to results obtained on each geometries with significantly different parametrisations (standard weekday and standard weekend day).

For the different experience we ran, for both aggregated indicators and time-series, it appears clearly that the effect of geometry is relatively neglectible regarding the role of parametrisation, what confirms the external consistence of the model regarding the study area.

Exploration of the parameter space A further step in the understanding of the behavior of the model, that can be considered as superficial in the overall process, but that is in fact implicitely crucial, is a grid exploration of the parameter space. We plotted surfaces for aggregated indicators as functions of all possible couples of parameters. The important fact is that we did not observe chaotic events, and especially for the mean-square-error surface, a quite continuous shape appeared, what suggested the use of a reduced calibration procedure, simplifying significantly the use of the model of simulation.

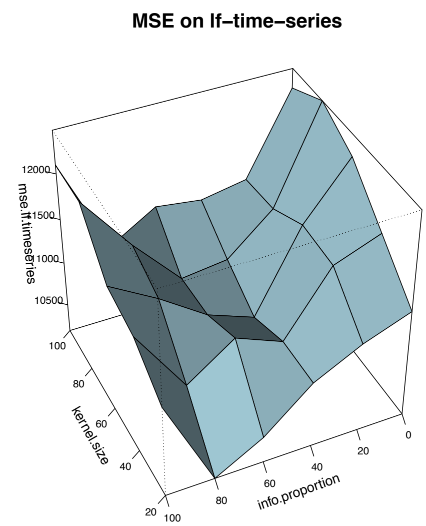
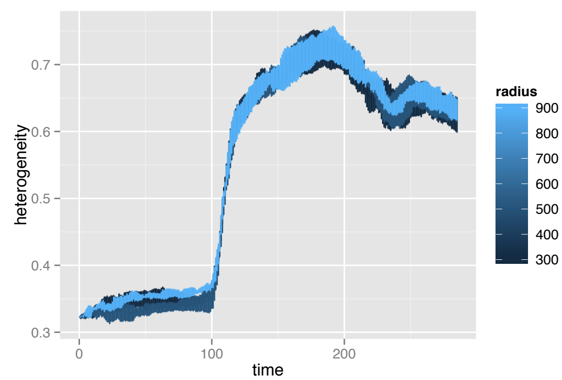


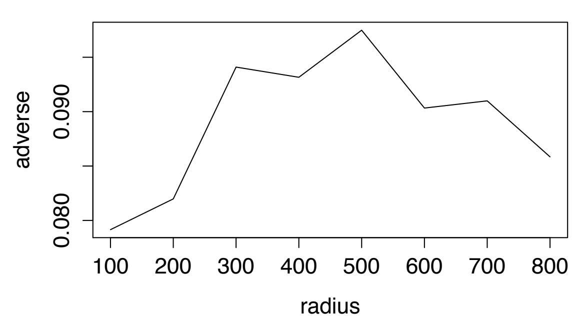
Fig. 4 | Simplified calibration procedure. We plot the surface of the mean-square error on time-series of load-factors as a function of the two parameters on which we want to calibrate. For visibility purpose, only one surface was represented out of the different obtained for different values of walking radius. The absolute minimum obtained for very large kernel has no sense since such value give quasi-uniform probabilities because of total recovering of Gaussian kernels. We take as best realisation the second minimum, which is located around a kernel size of 50 and a quantity of information fo 30%, which seem to be reasonable values afterwards.

Reduced calibration procedure Using experiments launched during the grid-exploration of the parameter space, we are able to assess or the regularity of some aggregated criteria, especially of the mean-square error on loads factors of stations. We calibrate on kernel size and quantity of information. For different values of the walking radius, the obtained area for the location of the minimum of the mean-square error stays quite the same for reasonnable values of the radius (300-600m). Fig. 4 shows an example of the surface used for the simplified calibration. We extract from that the values of around 50 for kernel size and 30 for information proportion. The most important is kernel size since we cannot have real proxy for that parameter. We use that values for the explorations of strategies in the following.

3.3 Investigation of user-based strategies



(a) Time series of heterogeneity indicators for different values of walking radius. Although small perturbations in the mean appear that could lead to misconclude to a positive effect of radius on heterogeneity. However, the error bars of all curves recover themselves, what leads to the non-significance of possible conclusion.



(b) Influence of walking radius on the quantity of adverse events. After 400m, we observe a decrease of the proportion. We can neglect values under 300-400m since these are smaller than the mean distance of a random point to a station, and would mean nothing, and conclude to a small effect of walking radius on adverse events proportion, but staying however quite prudent since the significance is quite weak.

Fig. 5 | Results on the influence of walking radius.

Influence of walking radius Taking for kernel-size and quantity of information roughly the values given by the calibration, we can test the influence of walking radius on the performance of the system. Note that we make a strong assumption, that is that the calibration stay valid for different values of the radius. As we stand previously, this stays true as soon as we stay in a reasonnable range of values (we obtained 300m to 600m) for the radius. Results are shown in figure 5.

Concerning the indicators evaluated on time-series, it is hard to have a significant conclusion since the small difference that one can observe between curves lies inside errors bars of all curves. For the quantity of adverse events, we see a decreasing of the indicator after a certain value (300m), what is significant if we consider that radius under that value are not realistic, since a random place in the city should be at least in mean over 300m from a bike station.

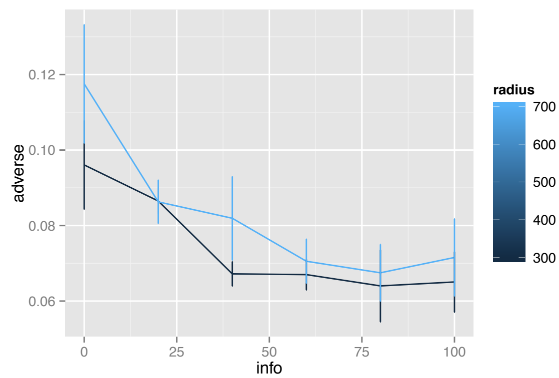
However, the results concerning the radius are not so concluding, what could be due to the presence of periodic negative feedbacks: when the mean distance between two stations is reached, repartitions concerns neighbor stations as expected, but the relation is not necessary positive, depending on the current status of the other station. A deeper understanding and exploration of the behavior of the model regarding radius should be the object of further work.

Influence of information For the quantity of information, we are on the contrary able to draw strong conclusions. Results from time-series are also not significant, but concerning aggregated indicators, we have a constant and regular decrease for each and for different values of the radius. We are able to quantify a critical value of the information for which most of the progress concerning adverse events is done, that is around 35%. We observe for this value an amelioration of 4% in the quantity of adverse events, that is interesting when rapported to the total number of bikers. Results are shown in Fig. 6.

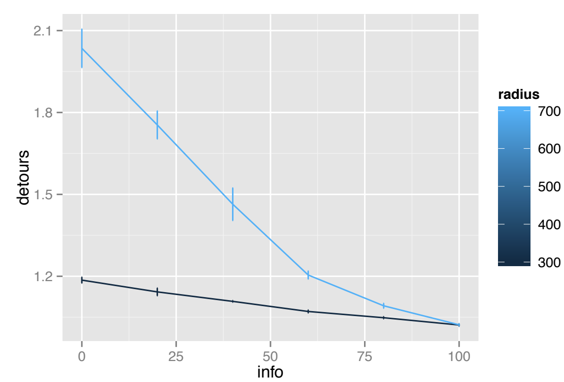
4 Discussion

Applicability of the results We have shown that increases of both walking radius and information quantity could have positive consequences on the level of service of the system, by reducing the overall number of adverse events and the quantity of detours especially in the case of the information as variable. However, we can question the possible applicability of the results. Concerning walking radius, first a deeper investigation would be needed for confirmation of the weak tendancy we observed, and secondly it appears that in reality, it should be infeasible to play on that parameter. The only way to approach that would be to inform users of the potential increase in the level of service if they are ready to make a little effort, but that is quite optimistic to think that they will apply systematically the changes, either because they are egoistic, because they won't think about it, or because they will have no time.

Concerning the information proportion, we cannot also force users to have information device (although a majority of population owns such a device, they won't necessarly install the needed software, especially if that one is not ergonomic). We should proceed indirectly, for example by increasing the ergonomy of the application. Information integrated in stations already exists but is not practical at all. That could be also a good way to have an influence on that parameter.



(a) Influence of proportion of information on adverse events for two different values of walking radius. We can conclude significantly that the information has a positive influence. Quantitavely, we extract the threshold of 35% that corresponds to the majority of decrease, that means that more is not necessarly needed.



(b) Influence of information on quantity of detours. Here also, the influence is positive. The effect is more significant for high values of walking radius.

Fig. 6 | Results on the influence of proportion of information.

Other possible strategies We have explored so far the role of behavior of users of the system. The object of further investigation could be the role of the "behavior" of stations. For example, one could fix rules to them, as close all parkings over a certain threshold of load-factor, or allow only departures or parkings in given configurations, etc. Such intelligent agents would surely bring new means to influence the overall system, but will also increase the level of complexity (in the sense of model complexity, see [32]), and therefore that extension should be considered very carefully (that is the reason why we did not integrate it in this first work).

Towards an online bottom up pilotage of the bikes sharing system Making the stations intelligent can imply making them communicate and behave as a self adapting system. If they give information to the user, heterogeneity in the nature and quantity of information provided could have strong impact on the overall system. That raises of course ethical issues since we are lead to ask if it is fair to give different quantities of information to different users. However, the exploration of such a self-adapting system, that would be equivalent to a bottom-up piloted system since the operator would have access to the parameters of the stations, could be strongly meaningful for a theoretical and practical point of view, beyond this particular sharing-bike problem. One could think of online adaptative algorithms for ruling the local behavior of the self-adapting system, such as ant algorithms ([33]), in which bikers would depose virtual pheromons when they visit a station (corresponding to their information on travel that is easy to obtain).

Conclusion

This work is a first step of a new bottom-up approach of bike-sharing systems. We have implemented, parametrised and calibrated a basic behavioral model and obtained interesting results for user-based strategies for an increase of the level of service.

Further work will consist in the extension to other types of agents such as stations, and in the study of possible bottom-up online algorithm for an optimal pilotage of the system.

Supplementary materials

S1 Report of statistical analysis of large set of data for VLib system, containing detailed procedure and protocol of primary data collection and detailed procedure of statistical treatment of the real time-series data, of used statistical models for inference of O, D fields and N_I, N_D random variables and of general interaction procedure between R and Netlogo.

S2 Global source code, including data collection scripts, statistical R code and NetLogo agent-based modeling code.

References

[1] Peter Midgley. The role of smart bike-sharing systems in urban mobility. *JOURNEYS*, 2:23–31, 2009.

[2] Paul DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4):41–56, 2009.

[3] Oliver O'Brien, James Cheshire, and Michael Batty. Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 2013.

[4] Zhili Liu, Xudong Jia, and Wen Cheng. Solving the last mile problem: Ensure the success of public bicycle system in beijing. *Procedia-Social and Behavioral Sciences*, 43:73–78, 2012.

[5] Xue Geng, Kai TIAN, Yu ZHANG, and Qing LI. Bike rental station planning and design in paris [j]. *Urban Transport of China*, 4:008, 2009.

[6] Jonathan Gifford and Arlington Campus. Will smart bikes succeed as public transportation in the united states? *Center for Urban Transportation Research*, 7(2):1, 2004.

[7] Pierre Borgnat, Patrice Abry, and Patrick Flandrin. Modélisation statistique cyclique des locations de vélo'v à lyon. In *XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2009.

[8] Pierre Borgnat, Eric Fleury, Céline Robardet, Antoine Scherrer, et al. Spatial analysis of dynamic movements of vélo'v, lyon's shared bicycle program. In *European Conference on Complex Systems 2009*, 2009.

[9] Pierre Borgnat, Patrice Abry, Patrick Flandrin, Jean-Baptiste Rouquier, et al. Studying lyon's vélo'v: a statistical cyclic model. In *European Conference on Complex Systems 2009*, 2009.

[10] Pierre Borgnat, Patrice Abry, Patrick Flandrin, Céline Robardet, Jean-Baptiste Rouquier, and Eric Fleury. Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems*, 14(03):415–438, 2011.

[11] Andreas Kaltenbrunner, Rodrigo Meza, Jens Grivolla, Joan Codina, and Rafael Banchs. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466, 2010.

[12] Jenn-Rong Lin, Ta-Hui Yang, and Yu-Chung Chang. A hub location inventory model for bicycle sharing system design: Formulation and solution. *Computers & Industrial Engineering*, 2011.

[13] Jenn-Rong Lin and Ta-Hui Yang. Strategic design of public bicycle sharing systems with service level constraints. *Transportation research part E: logistics and transportation review*, 47(2):284–294, 2011.

[14] Alvina GH Kek, Ruey Long Cheu, and Miaw Ling Chor. Relocation simulation model for multiple-station shared-use vehicle systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1986(1):81–88, 2006.

[15] Rahul Nair and Elise Miller-Hooks. Fleet management for vehicle sharing operations. *Transportation Science*, 45(4):524–540, 2011.

[16] Rahul Nair, Elise Miller-Hooks, Robert C Hampshire, and Ana Bušić. Large-scale vehicle sharing systems: Analysis of vélib'. *International Journal of Sustainable Transportation*, 7(1):85–106, 2013.

[17] Matthew Barth and Michael Todd. Simulation model performance analysis of a multiple station shared vehicle system. *Transportation Research Part C: Emerging Technologies*, 7(4):237–259, 1999.

[18] Matthew Barth, Michael Todd, and Lei Xue. User-based vehicle relocation techniques for multiple-station shared-use vehicle systems. *TRB Paper No. 04-4161*, 2004.

[19] Arnaud Banos, Annabelle Boffet-Mas, Sonia Chardonnel, Christophe Lang, Nicolas Marilleau, Thomas Thévenin, et al. Simuler la mobilité urbaine quotidienne: le projet miro. *Mobilités urbaines et risques des transports*, 2011.

[20] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences*, 20:514–523, 2011.

[21] U. Wilensky. Netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL., 1999.

[22] QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2009.

[23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[24] Jan C Thiele, Winfried Kurth, and Volker Grimm. Agent-based modelling: Tools for linking netlogo and r. *Journal of Artificial Societies and Social Simulation*, 15(3):8, 2012.

[25] Arnaud Banos. *Pour des pratiques de modélisation et de simulation libérées en Géographie et SHS*. PhD thesis, UMR CNRS Géographie-Cités, ISCPIF, Décembre 2013.

[26] Jonathan Bennett. *OpenStreetMap*. Packt Publishing, 2010.

[27] Alex Couture-Beil. rjson: Json for r. *R package version 0.2*, 13, 2013.

[28] Timothy H Keitt, Roger Bivand, Edzer Pebesma, and Barry Rowlingson. rgdal: bindings for the geospatial data abstraction library. *R package version 0.7-1, URL http://CRAN. R-project. org/package= rgdal*, 2011.

[29] Pablo Jensen, Jean-Baptiste Rouquier, Nicolas Ovtracht, and Céline Robardet. Characterizing the speed and paths of shared bicycle use in lyon. *Transportation research part D: transport and environment*, 15(8):522–524, 2010.

[30] T Warren Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857–1874, 2005.

[31] Alexandre B Tsybakov. Introduction to nonparametric estimation. (introduction à l'estimation non-paramétrique.). 2004.

[32] Franck Varenne, Marc Silberstein, et al. *Modéliser & simuler. Epistémologies et pratiques de la modélisation et de la simulation, tome 1.* 2013.

[33] Nicolas Monmarché. *Algorithmes de fourmis artificielles: applicationsa la classification eta l'optimisation*. PhD thesis, École Polytechnique, 2004.