

# Les apports de l'analyse prédictive des comportements humains à la création de villes plus agréables à vivre

Par Raphaël CHERRIER  
Fondateur de Qucit

Du fait d'une densité importante qui favorise les interactions et oblige à la mutualisation des ressources, les villes sont par excellence des lieux de concentration de la complexité. Cette complexité s'accompagne d'un certain nombre de défis à relever aussi bien dans le domaine de l'organisation des mobilités que dans celui de l'aménagement d'espaces publics agréables ouverts à l'ensemble des catégories d'usagers. Nous prendrons ici l'exemple des systèmes de vélos en libre-service pour illustrer les effets de réseau qui se développent dans un réseau dense, et les problématiques de congestion qui en découlent, avec les défis que cela pose en matière d'exploitation. Nous montrerons également comment l'analyse prédictive des comportements peut être utilisée pour fournir des solutions permettant de gérer les problématiques de congestion, ainsi que pour évaluer et, potentiellement, améliorer le confort dans les espaces publics.

## La complexité des villes

Les villes sont les lieux de concentration de la complexité. Celle-ci découle d'une forte densité de population, qui engendre une démultiplication des interactions humaines.

Vu du côté positif, cela permet de stimuler la créativité, le développement économique et culturel, et de réaliser des économies d'échelle en mutualisant les ressources.

Du côté négatif, cela provoque également plus de congestion, plus de criminalité, plus de pollution et des problèmes d'approvisionnement en ressources naturelles (nourriture, énergie, eau potable) et de gestion des déchets.

Il existe même en la matière une loi universelle <sup>(1)</sup> : à chaque fois que la population d'une ville double, toutes les valeurs dégagées par les interactions humaines augmentent de 15 %, tandis que les coûts d'infrastructure rapportés au nombre d'habitants diminuent eux aussi de 15 %.

## La mutualisation des ressources

Néanmoins, ces économies d'échelle ne peuvent être réalisées qu'en partageant les ressources.

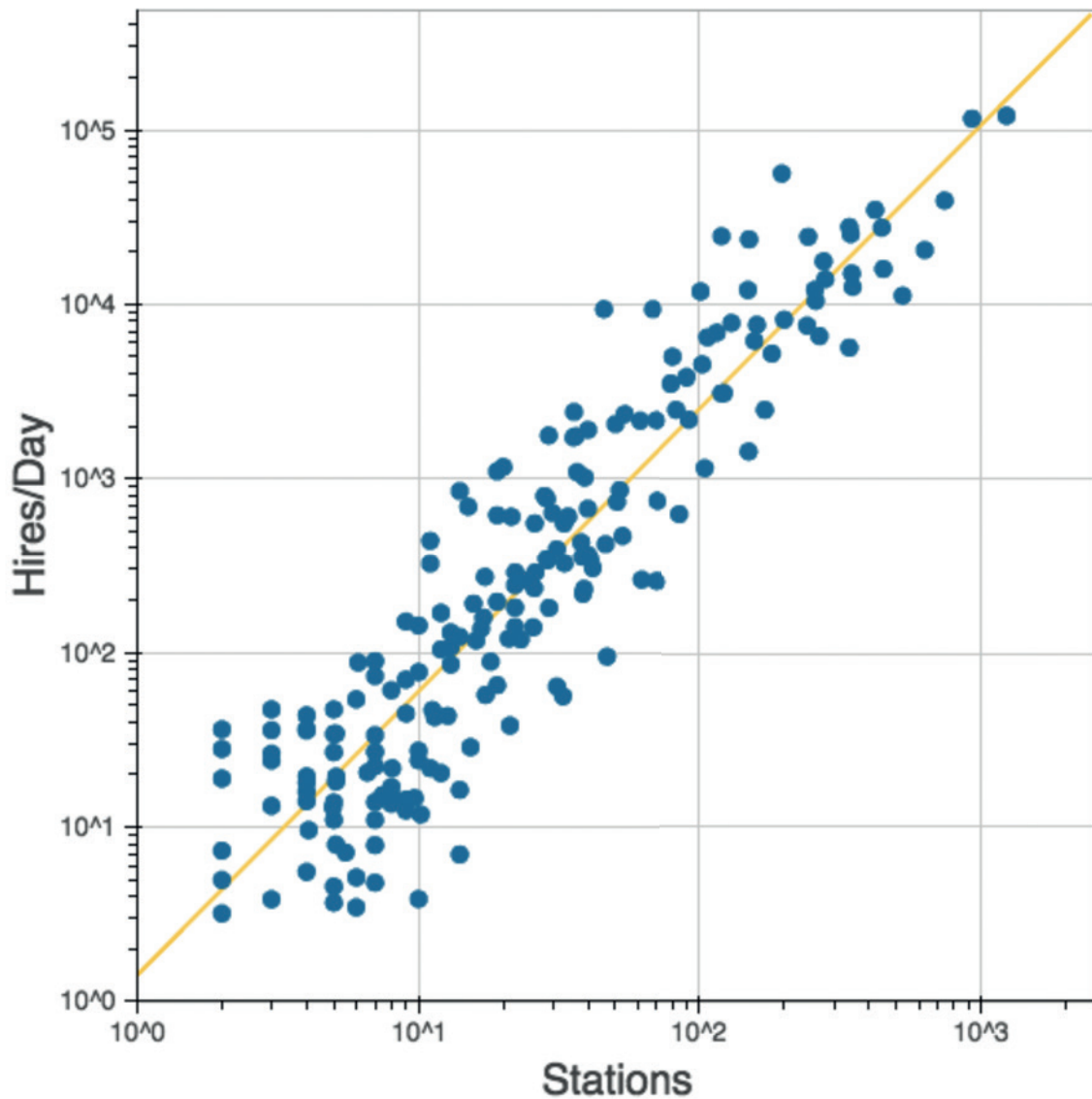
Ainsi, en ville, la plupart des habitants ne possèdent pas leur propre jardin et partagent les espaces publics pour

leurs loisirs ; les modes de transports sont eux aussi partagés (transports en commun, véhicules en libre-service, taxis). De même, les automobilistes partagent les places de stationnement en voirie, et la rue est partagée entre les voitures, les vélos et les piétons.

De manière moins visible, les réseaux d'approvisionnement (en eau, en électricité, en nourriture...) et la collecte des déchets, eux aussi mutualisés, requièrent d'importants efforts logistiques pour garantir à l'ensemble de la population l'accès à ces ressources vitales.

Par opposition à un modèle où chacun dispose en permanence d'un accès garanti aux ressources (ce qui est peut être le cas dans des zones moins densément peuplées), la ville est donc le lieu où un partage intelligent des ressources s'impose.

(1) BETTENCOURT (L.M.A.), LOBO (J.), HELBING (D.), KUHNERT (C.) & WEST (G.B.), "Growth, innovation, scaling, and the pace of life in cities", Proceedings of the National Academy of Sciences, 104 (17), 2007, pp. 7301-7306.



Analyse transactionnelle des taux moyens d'utilisation quotidiens des stations de VLS d'un échantillon de deux cents villes par nombre de stations.

### La start-up Qucit

Qucit (pour *Quantified Cities*) est une *start-up* française que nous avons créée en 2014. Notre mission est de fournir des modèles prédictifs de la mobilité, et plus généralement des comportements humains, basés sur l'analyse – grâce à des algorithmes d'intelligence artificielle – de l'ensemble des données disponibles sur les villes qui nous sollicitent et sur leurs habitants (données sur l'infrastructure urbaine, données sociodémographiques, données sur les événements, etc.).

Nous développons des applications concrètes à destination d'entreprises ou de collectivités agissant pour des villes plus durables, plus efficaces et plus agréables à vivre, et ce, dans des domaines variés liés aux problématiques de la ville intelligente (véhicules partagés, stationnement automobile, transports en commun ou encore réaménagement de l'espace urbain).

### Un exemple de service urbain mutualisé : les systèmes de vélos en libre-service (VLS)

Historiquement, les VLS ont vu le jour (sous leur forme actuelle) à Lyon, en 2005. Ils se sont depuis lors répandus rapidement à travers le monde : en 2016, près de 1 000 villes en sont équipées, totalisant plus de 1,5 million de vélos en circulation. Au niveau mondial, le marché de ce service (en croissance de 20 % par an) est estimé à 3 milliards de dollars <sup>(2)</sup>.

Ces systèmes offrent par ailleurs une illustration intéressante des problématiques que l'on peut rencontrer lorsqu'une ressource est mutualisée.

(2) Roland Berger (2014) : [https://www.rolandberger.com/en/Publications/pub\\_bike\\_sharing.html](https://www.rolandberger.com/en/Publications/pub_bike_sharing.html)

Par ailleurs, l'occupation des stations (le nombre des vélos disponibles) étant le plus souvent consultable en temps réel sur Internet, nous avons pu montrer qu'il est possible, en collectant et en analysant ces données, d'avoir une vision incroyablement précise du pouls d'une ville – tel que l'on peut le mesurer à travers le prisme des déplacements à vélo – et de recourir à l'analyse prédictive pour optimiser la conception et l'exploitation de ces systèmes.

### Effets d'échelle

De nombreux systèmes de VLS de petite taille peinent à trouver une rentabilité, leurs taux d'utilisation restant très faibles.

Mais en augmentant la taille du système, on crée un effet de réseau qui démultiplie les possibilités de déplacements (le nombre des trajets possibles étant proportionnel au carré du nombre de stations du réseau).

L'analyse de plus de 200 systèmes de VLS d'une trentaine de pays a révélé qu'en moyenne, à chaque fois que la taille du système double, le nombre des trajets effectués est multiplié par 3<sup>(3)</sup>. En particulier, le taux de rotation des vélos et l'utilisation des infrastructures sont ainsi augmentés de 50 %.

### Pénurie et congestion

Lorsque l'utilisation du système est importante, ce qui arrive à partir d'une certaine taille critique (typiquement une centaine de stations) avec l'effet de réseau, un certain nombre de stations se retrouvent en situation de pénurie de vélos disponibles, ou au contraire congestionnées (sans emplacement disponible pour pouvoir déposer le vélo emprunté).

Cela s'explique tout d'abord, par le fait que **les flux ne sont pas parfaitement symétriques**. Par exemple, dans la plupart des villes, les lieux de sorties sont situés au centre-ville : certains habitants s'y rendent en transport en commun en début de soirée et repartent en VLS, générant ainsi un flux net de vélos orienté du centre-ville vers la périphérie (les différences d'altitude entre le point de départ et celui d'arrivée constituent elles aussi une cause majeure de dissymétrie).

Cela s'explique aussi par le fait que **les flux de VLS sont pendulaires et dépendent de variables « structurelles » qui caractérisent la ville** (les infrastructures de transport, la démographie, l'activité économique, etc.). On observe ainsi les tendances de fond suivantes : les zones résidentielles se vident le matin au profit des zones d'activité et des zones d'emploi ; les lieux de sorties se remplissent en début de soirée et se vident en fin de soirée ; les lieux où se situent des restaurants sont fortement fréquentés le midi ; les gares subissent également des flux pendulaires liés aux voyageurs quittant la ville et aux habitants de la périphérie venant travailler en ville ; etc.

Enfin, les flux de VLS dépendent également de **variables « dynamiques »**. Ils varient, bien sûr, selon le jour de la semaine ou en fonction du calendrier des jours fériés ou des vacances scolaires. Mais ils vont également être modifiés par l'occurrence d'un concert ou d'un événement

sportif, par une panne ou une grève dans les transports en commun, et, dans le cas des vélos en libre-service, ils sont très fortement influencés par la météo. Mais dans la pratique, un grand nombre d'autres variables vont avoir un impact plus ou moins important (comme la pollution de l'air, ou même le niveau de stress ressenti par la population suite, par exemple, à un attentat).

## Pourquoi prédire les comportements humains ?

Ces problèmes de disponibilité des vélos sont la principale source de mécontentement des usagers des VLS (avec les problèmes liés aux vélos endommagés, ou aux bornes défectueuses) et contribuent à donner l'image d'un service peu fiable.

### Fournir de l'information fiable

Pour maintenir la confiance des usagers et faire des VLS de véritables moyens de transport intégrés avec les transports en commun, il est nécessaire de mettre au point des systèmes d'information adaptés prenant en compte les contraintes de disponibilité.

Par exemple, pour pouvoir proposer un itinéraire mixte constitué d'un tronçon en transport en commun suivi d'un tronçon en VLS, il est indispensable de pouvoir garantir à l'utilisateur qu'il disposera bien d'un vélo au point de rupture modale et d'une place disponible à l'arrivée. Or, aux heures de pic d'activité, les stations de VLS peuvent se remplir ou se vider extrêmement rapidement, et il s'avère que l'information en temps réel n'est dès lors plus suffisante.

Chez Qucit, nous avons développé le premier moteur de recherche prédictif au monde pour les VLS, capable de prédire les disponibilités des vélos et des points d'attache en stations jusqu'à 12 heures à l'avance<sup>(4)</sup>.

### Participer à l'étalement de la congestion

Le deuxième intérêt de l'information prédictive est qu'elle permet d'étalement la congestion en encourageant les voyageurs à se diriger vers les stations où la probabilité de trouver un vélo (ou une borne) disponible est la plus élevée.

Nous éditons à cette fin BikePredict, une application mobile disponible dans plus de 200 villes d'Europe : il s'agit d'un assistant personnel pour l'utilisation des VLS qui fonctionne grâce à des algorithmes prédictifs.

BikePredict ainsi que les calculateurs d'itinéraires multimodaux basés sur nos algorithmes prédictifs participent au rééquilibrage des réseaux de VLS sur de courtes échelles de distance (environ 500 mètres au maximum) correspondant à la distance supplémentaire que l'utilisateur d'un VLS est prêt à parcourir à pied.

(3) Qucit (2015) : <http://www.qucit.com/2015/10/15/velos-en-libre-service-limportance-de-la-taille-du-reseau/>

(4) Avec une fiabilité à une heure de 97 % pour le Vélib', et supérieure pour la grande majorité des autres systèmes de VLS.

## Optimiser l'exploitation

La méthode utilisée pour rééquilibrer les systèmes de VLS consiste à déplacer des vélos – le plus souvent grâce à des camionnettes – des stations pleines vers les stations vides ; le déplacement des vélos, dont le nombre se situe entre 3 et 5 % du nombre des emprunts réalisés par les usagers, permet d'atteindre un taux moyen de disponibilité de l'ordre de 85 à 90 % (dans la plupart des grandes villes françaises).

En anticipant les flux d'utilisation des VLS plusieurs heures à l'avance, les modèles prédictifs permettent de développer des stratégies de rééquilibrage optimal des stations, de diminuer le nombre des kilomètres parcourus par les camionnettes servant à la régulation et les coûts d'exploitation afférents, d'augmenter la disponibilité des stations, et finalement d'augmenter l'utilisation du système et la rentabilisation des investissements (installation des stations, achats de vélos...) <sup>(5)</sup>.

## Planifier les investissements d'infrastructure

En analysant le lien entre les données structurelles de la ville considérée et l'utilisation que font les habitants usagers de ses stations de VLS, il est possible de prévoir leur taux d'utilisation et de dimensionner ainsi de manière optimale le système - tout en prenant en compte l'impact ultérieur de l'utilisation effective sur les opérations de réassort.

## Approche quantitative

### Les comportements humains

La première catégorie de comportements auxquels nous nous intéresserons concerne ceux qui décrivent l'activité humaine, par exemple :

- la mobilité, en tenant compte des différents modes de transport utilisés (marche, vélo, vélo en libre-service, trafic automobile, transports en commun, stationnement automobile),
- les comportements d'achat, les consommations d'eau, d'électricité...
- la criminalité, la fraude au stationnement, les dégradations, etc.

La seconde catégorie concerne tout ce qui a trait au **ressenti**, aux **émotions** et aux **opinions**, par exemple :

- les sentiments de confort, de stress ou de sécurité ressentis dans un espace public,
- l'opinion sur l'esthétique et/ou sur la propreté d'un lieu.

Traditionnellement, l'activité humaine est comptabilisée ou mesurée par des capteurs. Les opinions le sont, quant à elles, au moyen de sondages.

Néanmoins, les sondages sont également utilisés pour mesurer l'activité humaine, lorsque la mesure directe de celle-ci est trop complexe. Ce sont par exemple les enquêtes origine-destination ou les enquêtes ménage-déplacement réalisées dans les transports.

### Prédire les comportements humains

Notre vision est que la multiplication des données numériques et les avancées de l'intelligence artificielle per-

mettent aujourd'hui d'aborder la prédiction des comportements humains par le biais d'une approche quantitative qui s'affranchit en grande partie du caractère subjectif de l'observation humaine et de la modélisation par un expert.

Pour cela, le premier élément indispensable est de **disposer de mesures précises des comportements humains et des contextes dans lesquels ils interviennent**.

Par comparaison avec la théorisation des lois de la Nature opérée par les physiciens, la difficulté principale est qu'il est quasi impossible de réaliser des expériences dans un contexte contrôlé et simplifié afin de pouvoir isoler l'influence d'une variable sur les comportements des individus.

Pouvoir mettre en évidence l'influence de chacune des variables prise isolément demande alors de disposer :

- d'une description aussi exhaustive que possible du contexte (environnement urbain, mais aussi caractéristiques de l'individu) dans lequel les comportements humains se produisent,
- d'un grand nombre d'observations,
- de la capacité de traiter cette énorme quantité d'informations hétérogènes.

L'autre élément indispensable à la démarche de théorisation est donc la possibilité de **disposer des bons outils de modélisation**.

Les humains perçoivent le monde à travers des concepts complexes qu'il est parfois difficile de mettre en évidence parmi une multitude de données brutes. Par exemple, comment déterminer, à partir du plan d'une ville, même extrêmement détaillé, qu'un lieu va être perçu comme agréable ?

### L'évolution des dispositifs de mesure

Nous partons du principe que tout type de comportement humain est prédictible, la qualité de la prédiction étant directement fonction de la quantité et de la variété des données contextuelles disponibles et de notre capacité à traiter ces dernières de manière algorithmique.

La **numérisation croissante des villes et des modes de communication** révolutionne la manière de mesurer et de prédire les comportements humains <sup>(6)</sup>. Des données en nombre exponentiellement croissant sont produites par :

- des capteurs disposés dans l'espace urbain qui mesurent des paramètres variés, comme le bruit, la pollution, la mobilité, l'occupation des places de stationnement... <sup>(7)</sup>,

(5) Nous avons montré qu'il est possible d'obtenir une amélioration de l'ordre de 20 % du taux de disponibilité des stations en utilisant l'analyse prédictive pour optimiser les stratégies de réassort : [http://www.qucit.com/wp-content/uploads/sites/3/2016/05/201510\\_Final-Paper\\_Qucit\\_ITSBikes.pdf](http://www.qucit.com/wp-content/uploads/sites/3/2016/05/201510_Final-Paper_Qucit_ITSBikes.pdf)

(6) D'ici à 2030, la quantité des données produites dans les villes devrait être multipliée par 100 (en supposant que le rythme actuel d'un doublement tous les 2 ans perdure).

(7) Selon une étude de Gartner (2015), le nombre d'appareils connectés à Internet sera de 21 milliards en 2020 (à comparer à 6,4 milliards en 2016).



- les *smartphones*, qui sont équipés de puces GPS, d'accéléromètres, de capteurs de pression, de luminosité ou de bruit...
- les communautés d'utilisateurs d'applications mobiles,
- enfin, les réseaux sociaux.

Ces données permettent de mesurer un grand nombre de phénomènes de manière plus précise, moins coûteuse, et ce, avec une latence plus réduite que ne le permettent les méthodes traditionnelles <sup>(8)</sup>.

### L'intelligence artificielle

Les progrès en intelligence artificielle permettent :

- de traiter automatiquement et pour un faible coût de grandes quantités de données hétérogènes,
- d'apprendre des concepts complexes à partir de données qu'il aurait été impossible de programmer explicitement.

Par ailleurs, nous disposons de **bibliothèques logicielles d'apprentissage automatique *Open Source*** <sup>(9)</sup> de plus en plus performantes, ainsi que de la puissance de calcul nécessaire pour les faire fonctionner sans même avoir besoin d'investir dans l'achat de serveurs.

### Applications

Chez Qucit, notre mission de quantifier et de mathématiser les comportements humains dans les zones denses nous a conduits jusqu'à présent à appliquer à un certain nombre de problématiques variées le formalisme que nous avons décrit plus haut.

Notre projet **CityPark** consiste à prédire le temps moyen nécessaire à un automobiliste pour trouver une place de stationnement en voirie en fonction de son lieu de destination et de son heure d'arrivée. En fonction de ses préférences personnelles, l'automobiliste est guidé vers une solution optimale : stationnement en voirie au plus près de sa destination (ou dans un quartier proche, mais moins congestionné), stationnement en ouvrage ou dans un parking partagé.

Notre conviction est qu'en informant les automobilistes de la congestion des différentes zones de la ville au moyen de données concrètes, notamment des temps de trajet, il est possible de générer des modifications de comportement de nature à fluidifier le stationnement en ville, pour le bénéfice de tous.

Techniquement, CityPark utilise des données issues de l'analyse des traces GPS et d'accélérométrie que nous collectons (pendant au maximum une heure) auprès des utilisateurs qui nous y autorisent. Ces données servant à calibrer des modèles contextuels.

Un autre de nos projets, réalisé en partenariat avec la Mairie de Paris et Cisco <sup>(10)</sup>, consiste à réaliser un **indicateur quantitatif du confort ressenti par les piétons**. L'indicateur que nous avons mis au point inclut six dimensions (confort, stress, sécurité, désorientation, avis sur l'esthétique et la propreté du lieu).

L'objectif était d'accompagner les urbanistes de la Mairie de Paris en leur proposant des outils pour quantifier l'impact des réaménagements prévus Place de la Nation sur la réappropriation de l'espace par les piétons et sur la modification des flux et de l'indicateur quantitatif de confort mis au point.

Dans ce cas, les données de calibrage ont été obtenues par une enquête de terrain réalisée auprès de 1 300 passants entre le 29 mars et le 3 avril 2016.

La disponibilité de l'indice de confort mis au point par Qucit (Confort Paris) ouvre la porte à l'expérimentation : la gestion des espaces publics peut devenir un processus itératif et quantitatif. Ainsi des solutions peuvent être trouvées en utilisant l'analyse prédictive pour maximiser le bien-être des citoyens.



Modèle contextuel des niveaux de stress ressentis par les piétons Place de la Nation, dans l'est de Paris.

(8) Un exemple est l'application Waze, dont les données de trafic et les déclarations d'accidents révolutionnent la manière de gérer le trafic, une application qui pourrait même permettre d'organiser plus rapidement les secours en cas d'accident : <https://www.washingtonpost.com/news/tripping/wp/2016/05/11/can-waze-speed-up-emergency-response-times-perhaps-as-data-sharing-becomes-two-way-street/>

(9) <https://www.tensorflow.org/>

(10) En 2016, Cisco et la Ville de Paris se sont engagés dans un partenariat par lequel la Ville accorde l'accès à des espaces publics autour de la Place de la Nation, pour permettre à Cisco d'installer des capteurs destinés à mesurer les mouvements des passants ou des voitures, la pollution, le bruit, etc. La Place de la Nation sert ainsi de pilote pour un aménagement des places parisiennes prévu dans le cadre du programme « Réinventons nos places » : <http://www.paris.fr/services-et-infos-pratiques/urbanisme-et-architecture/projets-urbains-et-architecturaux/reinventons-nos-places-2540>