

Introduction à un domaine de recherche: Étude sur l'optimisation des modes de collecte des déchets ménagers

Lucie Laurence - 20 mai 2021

Table des matières

1	Présentation du sujet	1
2	Modélisation mathématique	3
3	Cas d'un seul noeud	5
4	Simulation	7
5	Résultat de champ moyen	9
6	Conclusion et ouvertures sur la suite	9

1 Présentation du sujet

La collecte des déchets est un enjeu de taille dans nos sociétés actuelles. Aussi bien pour des raisons économiques qu'écologiques, les établissements chargés de s'en occuper ont tout intérêt à optimiser leur mode de fonctionnement. C'est ce qu'a souhaité faire Vallée Sud-Grand Paris (VSGP), un établissement territorial d'Ile de France réunissant 11 communes, 390 000 habitants dans 174 000 foyers. La collecte concerne 130 000 tonnes de déchets ménagers par an, et un budget annuel de 20 millions d'euros.

Soucieux de réduire leur impact écologique et leur budget annuel, Vallée Sud-Grand Paris a imaginé une nouvelle méthode de collecte qu'ils pourraient mettre en place : la collecte des bacs à la demande du particulier. Cette méthode innovante n'ayant jamais été testée à grande échelle, ils ont fait appel au département de mathématiques de l'ENS pour tenter d'obtenir une réponse théorique sur la possibilité de mettre en place ce système.

J'ai choisi de faire mon exposé sur cette modélisation, et non sur mon futur sujet de thèse car il me paraît important de présenter un sujet de mathématiques appliquées, au service de l'entreprise et d'un problème concret.

L'étude a pour but de comparer deux modes de collecte : le modèle régulier et le modèle majoritaire. Dans le cas du *modèle régulier*, les poubelles sont vidées périodiquement, tous les r jours, et ce quel que soit leur contenu. C'est le modèle qui est actuellement en place. Dans le *modèle majoritaire*, les poubelles ne sont vidées que quand elles sont pleines. Plus précisément, quand un bac dépasse un seuil de remplissage s , son propriétaire demande à ce qu'il soit vidé. Pour faire passer une benne qui ramassera les bacs pleins, il faut qu'un minimum d'appel (k , seuil de ramassage) ait été effectué dans une même rue. C'est un système de collecte "à la demande", et le modèle que Vallée Sud-Grand Paris souhaiterait mettre en place.

L'objectif de cette étude est de montrer la viabilité du modèle majoritaire (les bacs ne débordent pas ? le temps de collecte sera-t-il raisonnable ?), ainsi que de comparer ses coûts à ceux du modèle régulier. Pour cela, nous avons réfléchi majoritairement sur trois points :

- la *modélisation* mathématique des deux modèles.
- la mise en évidence d'*indicateurs* qui caractériseront l'efficacité des collectes.
- la *comparaison* de ces deux modèles, à l'aide des indicateurs.

Évidemment, l'étude effectuée lors de mon stage ne prétendait pas donner une conclusion définitive sur le choix de la collecte à utiliser. Elle a cependant permis de mettre en évidence certains phénomènes qui seront intéressants à considérer lors de discussions ultérieures.

Les travaux antérieurs

Le sujet, abordé de cette manière, n'avait jamais été étudié à ma connaissance.

Des modélisations ont déjà été faites pour étudier le trajet de collecte optimal. Voir par exemple [1]. Dans ces modélisations, c'est le trajet de la benne (camion de ramassage) qui est optimisé, en fonction de tous les bacs à collecter. Dans ces modèles la quantité de déchets dans chaque noeud est souvent déterministe, et l'apport journalier est constant. Une telle modélisation rend la collecte régulière optimale, et ne pourra pas permettre une bonne comparaison des deux modèles.

Cette absence de références antérieures a fait la difficulté de ce sujet, puisque tout était encore à construire, mais nous a également permis une grande liberté.

Le même sujet a été proposé par Bertrand Maury lors de son atelier *Maths et entreprise* de l'ENS cette année, et les trois élèves ayant travaillé sur le sujet sont partis dans des directions très différentes.

Présentation de l'étude

L'étude a été effectuée sur plusieurs niveaux. Il a d'abord fallu comprendre le fonctionnement des collectes en s'entretenant avec Vallée Sud-Grand Paris, qui nous a fourni un certain nombre de documents tels que des rapports d'activités. Ceux-ci ont permis de faire une modélisation fidèle du problème. Nous avons commencé par étudier le comportement d'un seul bac avec les deux méthodes de collecte, puis nous nous sommes placés à l'échelle d'un secteur de collecte pour prendre en compte plusieurs foyers. Pour cela, nous avons d'abord fait des simulations, qui ont découlé sur une conjecture.

Vocabulaire technique : On appelle **bac** le conteneur qui permet à chaque foyer de stocker ses poubelles avant la **collecte**, moment où les bacs sont vidés par la **benne** (camion de poubelle). La **capacité** d'un bac est la quantité totale de poubelles qu'il peut contenir. Enfin, les foyers (détenant chacun un bac) sont appelés **noeuds** dans la suite.

La modélisation mathématique : Nous avons modélisé la production de déchets journalière par habitant par des variables aléatoires (variables de remplissage), et étudié les deux processus représentant le remplissage des bacs au temps t selon la politique de ramassage choisie. Nous nous sommes ensuite concentrés sur le modèle majoritaire avec un seul noeud : nous avons montré l'irréductibilité et la récurrence du processus, et en utilisant plusieurs distributions différentes pour les variables de remplissage, nous avons pu calculer des données-clefs telles que les mesures invariantes de probabilité, la distribution des noeuds aux jours de collecte, et le temps entre deux collectes.

La simulation : La partie simulation nous a confronté à une réflexion sur les données et ordres de grandeurs à utiliser pour simuler au mieux le problème.

Après avoir simulé le remplissage des bacs et les collectes selon les deux politiques, il a fallu comprendre comment exploiter les résultats des simulations, en dégagant les *indicateurs* pertinents.

Chacun de ces indicateurs met en jeu des ordres de grandeurs que l'on doit déterminer en observant les données réelles. Nous verrons que certains de ces ordres de grandeurs peuvent fausser les résultats s'ils ne sont pas calculés assez précisément.

Ensuite, nous avons utilisé les résultats de la collecte régulière pour comprendre dans quels scénarios la collecte majoritaire peut être mise en place : par exemple, la durée d'une collecte est de 7h dans le modèle régulier, 7h qui est la limite juridique autorisée pour une durée de collecte. Pour que la collecte majoritaire soit faisable, il faut que sa durée soit au plus de 7h, et donc le temps de collecte ne doit pas dépasser celui de la collecte régulière.

Enfin, nous avons interprété les résultats et cherché à comprendre la façon d'optimiser les paramètres du modèle.

2 Modélisation mathématique

Nous avons modélisé les deux politiques de ramassage avec des processus.

La première chose à modéliser est le remplissage quotidien des bacs. Nous avons regardé n foyers avec une production quotidienne *aléatoire* de déchets, qui ne dépend ni du jour ni du foyer concerné. C'est ce qu'on appelle un modèle *homogène en temps et en espace*. On note $\alpha_{i,t}$ les déchets produits par le noeud i au temps t . Les variables de remplissages $(\alpha_{i,t})$ sont donc indépendants et identiquement distribués.

Par ailleurs, on a considéré qu'un bac (récipient stockant les poubelles par foyer) a une capacité c . Nous avons étudié le remplissage des bacs au cours du temps, modélisé par un processus $(X(t))$.

On peut d'ores et déjà définir le processus associé au modèle régulier de fréquence de ramassage r :

Définition 2.0.1. Soit $n \geq 1$. Un processus aléatoire $X(t) = (X_i(t))_{1 \leq i \leq n}$ à valeurs dans \mathbb{N}^n est un **processus régulier** associé aux **variables de remplissages** $(\alpha_{i,t})$ et de **fréquence de ramassage** r s'il vérifie :

1. Les $(\alpha_{i,t})$ sont i.i.d à valeurs dans \mathbb{N}
2. r est un entier strictement positif
3. Pour tout $1 \leq i \leq n$, on a :

$$X_i(t) = 0 \text{ si } t \text{ est multiple de } r \quad (1)$$

$$X_i(t) = X_i(t-1) + \alpha_{i,t} \text{ sinon} \quad (2)$$

On notera que si t est un multiple de r , la somme est vide, et $X(t) = 0$ (tous les noeuds sont vidés).

Pour définir le processus suivant le modèle majoritaire, on doit introduire un seuil à partir duquel les propriétaires des bacs demanderont le ramassage. Un des impératifs émis par Vallée Sud-Grand Paris lors de nos entretiens pour préciser le modèle étudié était la nécessité d'éviter à tout pris le débordement des bacs. Pour ce faire, nous avons décidé de fixer le seuil d'appel s en dessous de la capacité du bac c .

Par ailleurs, nous introduisons un autre seuil k de ramassage, qui correspond au nombre minimum d'appels nécessaires avant de faire passer une benne pour collecter les bacs. On obtient finalement la définition suivante :

Définition 2.0.2. Soit $n \geq 1$. Un processus aléatoire $X(t) = (X_i(t))_{1 \leq i \leq n}$ à valeurs dans \mathbb{R}^n est un **processus majoritaire** associé aux **variables de remplissages** $(\alpha_{i,t})$, au **seuil de remplissage** s et au **seuil de ramassage** k s'il vérifie :

1. Les $(\alpha_{i,t})$ sont des variables aléatoires i.i.d à valeurs dans \mathbb{N}
2. s et k sont des réels strictement positifs
3. En posant

$$H_s(t) = \sum_{i=1}^n \mathbb{1}_{\{X_i \geq s\}} \quad (3)$$

pour tout $1 \leq i \leq n$, on a :

$$X_i(0) = 0 \tag{4}$$

$$X_i(t) = 0 \text{ si } H_s(t-1) \geq k \text{ et } X_i(t-1) > s \tag{5}$$

$$X_i(t) = X_i(t-1) + \alpha_{i,t} \text{ sinon} \tag{6}$$

On remarque que le poids du bac i revient à zéro si le bac est plein **et** qu'il y a assez de bacs pleins pour faire passer une benne.

3 Cas d'un seul noeud

Pour commencer, nous avons étudié le cas de $n = 1$ (un seul bac mis en jeux). On prend (α_i) une suite de VA à valeurs dans \mathbb{N} , indépendantes et identiquement distribuées, représentant la production journalière de déchets. $(X^r(t))$ et $(X^m(t))$ sont les processus décrivant l'évolution du remplissage du bac observé, dans le modèle régulier (ramassage tous les r jours) et le cas majoritaire (ramassage quand $X(t) \geq s$). On suppose par ailleurs que $0 < \mathbb{P}(\alpha_1 = 0) < 1$.

Les deux processus ainsi définis sont récurrents et positifs (quelle que soit la loi de α). Ils sont en effet irréductibles, et le retour en zéro a lieu le lendemain du dépassement du seuil s . Comme $\alpha \neq 0$, ce temps est d'espérance finie.

Le processus associé au modèle régulier n'est pas très intéressant : on calcule facilement sa distribution au ramassage (somme de r variables indépendantes et identiquement distribuées) , et son temps de ramassage est fixe (égal à r).

On peut tout de même noter que la capacité c des bacs est déterminée grâce à ce processus : on veut qu'au ramassage la probabilité que le bac déborde soit inférieure à $\mu = 1/20$. c est fixé par le μ -quantile de la variable $X^r(r)$ (qui dépend de la loi des α).

Désormais, nous nous intéressons au processus du modèle majoritaire.

La modélisation a pour but de récupérer des données-clefs sur notre modèle. Ici, on étudie le temps de ramassage et la distribution au ramassage du processus $(X^m(t))$.

On note dans la suite $\tau_s = \inf\{t \geq 0 : X^m(t) \geq s\}$.

On voudrait déterminer les distributions de τ_s et de $X^m(\tau_s)$.

On peut relier la distribution de $X^m(\tau_s)$ à la mesure invariante de $(X^m(t))$, on est donc ramené à un calcul de mesure invariante.

Proposition 3.0.1. *On note $p_0 = \mathbb{P}(\alpha_1 = 0)$, et m la mesure invariante qui vaut 1 en 0. Soit m' la restriction de m à $[s, \infty[$. Alors $X^m(\tau_s)$ a pour loi*

$$\frac{m'}{1 - p_0}$$

On peut appliquer cette proposition pour des distributions de α afin d'avoir une expression explicite de la distribution du bac au ramassage.

Pour la distribution du temps d'arrêt, on a également obtenu des formules explicites. Voici par exemple les résultats obtenus pour des variables de remplissage suivant une loi géométrique de paramètre p ($\mathbb{P}(\alpha_1 = k) = qp^k$, $0 < q = 1 - p$) :

Proposition 3.0.2. Si $(X^m(t))$ est le processus représentant un bac dans le modèle majoritaire, avec seuil s et variables de remplissage géométrique de paramètre p , en notant $\tau = \tau_s$ le temps de ramassage, on a :

$$\mathbb{P}(\tau = t) = \binom{t-1}{s-1} p^s q^{t-s} \delta_{t \geq s} \quad (7)$$

et $X^m(\tau) - s$ suit une loi géométrique de paramètre p .

Interprétation des résultats

Ces données sont à comparer avec les distributions dans le modèle régulier.

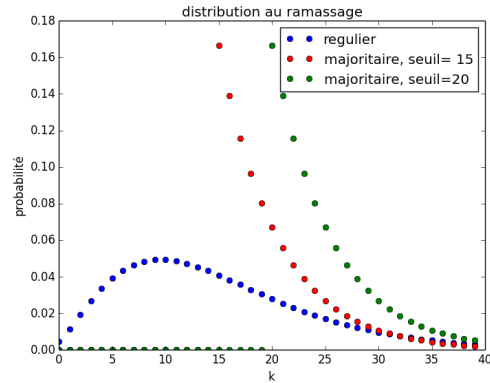


FIGURE 1 – Distribution des bacs au ramassage, cas géométrique

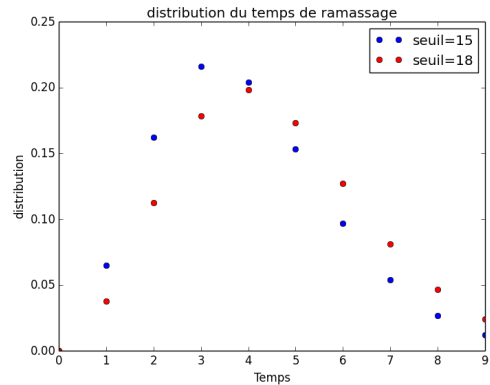


FIGURE 2 – Distribution du temps de ramassage dans le cas géométrique

Pour la représentation graphique, on a utilisé des variables de remplissage de loi géométrique de paramètre 5, on a posé $r = 3$ pour le modèle régulier. Nous

observons que le bac est bien plus plein au ramassage dans le modèle majoritaire (ce qui était prévisible). Par ailleurs, le temps de ramassage est en moyenne plus grand que 3 dans le modèle majoritaire, la benne passera donc moins souvent.

Idée de preuve

La preuve se fait en étudiant les propriétés du processus. On détermine la mesure invariante, et on la relie à la distribution de $X(\tau)$. Pour cela, nous avons dû introduire des processus intermédiaires qui simplifiaient les calculs, comme le processus réduit qui est strictement croissant jusqu'au ramassage.

4 Simulation

Ce premier modèle à un noeud n'est évidemment pas assez représentatif de la réalité pour faire obtenir des résultats concluants. Malheureusement, tels que les résultats ont été obtenus dans le cas $n = 1$, ils ne sont pas généralisables à un nombre plus importants de noeuds.

Pour essayer de mieux comprendre le système que nous étudions, nous avons effectué des simulations. Cela a permis de prendre en compte beaucoup plus de facteurs, et d'observer l'évolution de données clefs, au delà du remplissage des bacs.

Nous avons commencé par simuler le remplissage des bacs, avec des variables de remplissage suivant une loi de Poisson, nous avons ajouté les politiques de ramassage au programme. Ensuite, nous avons dû trouver les valeurs qui permettraient de comparer les modèles.

Variables, indicateurs et ordres de grandeurs

Le principal défi de la simulation a été de fixer toutes les variables et indicateurs à observer.

Parmi les variables, certaines (comme le nombre de noeuds observés) étaient à fixer en amont, tandis que d'autres étaient à optimiser (la capacité du bac c , fixé par l'étude d'un seul bac dans le cas régulier, les seuils s et k).

Les *indicateurs*, i.e les fonctions qui dépendent des paramètres à optimiser, sont les suivants :

- le nombre de kilomètres effectués par collectes
- la durée d'une collecte
- le nombre de bacs qui débordent au moment de la collecte
- la quantité de déchets collectée par heure
- la quantité de déchets collectée par kilomètre

Cette partie de la modélisation a dû se faire en lien étroit avec l'établissement VSGP, via Suzanne Rabaud qui a été mon interlocutrice tout au long du stage.

Ces indicateurs sont dépendants d'ordres de grandeurs qui ont dû être récupérés parmi les données de VSGP. Par exemple, le nombre de kilomètres effectués par la benne doit comprendre les trajets pour la vider, donc dépend de la taille de la benne, et de la distance aux centres de traitements des déchets. De

même, le temps de collecte dépend des vitesses du camion, et du temps nécessaire pour vider un bac. Certaines données étaient plus ou moins compliquées à déterminer, nous avons dû faire quelques études de terrain pour récupérer les données manquantes. Par exemple, des équipes de Vallée Sud-Grand Paris ont suivi des bennes de collecte pour essayer d'estimer le temps nécessaire pour vider un bac, et la vitesse moyenne du camion pendant la collecte.

Résultats obtenus

Nous illustrerons ici les conclusions auxquelles nous avons pu aboutir en observant les résultats de simulation sur la distance parcourue lors d'une collecte.

Sur ces graphiques (figure 3) sont représentés les kilomètres moyens effectués en une collecte, en fonction du choix du seuil s à $k = 50$ fixé, et en fonction de k à $s = 18$ fixé.

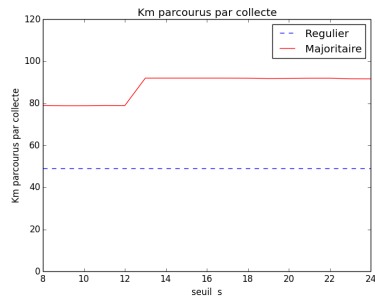


FIGURE 3 – Kilomètres parcourus en une collecte en fonction du seuil s

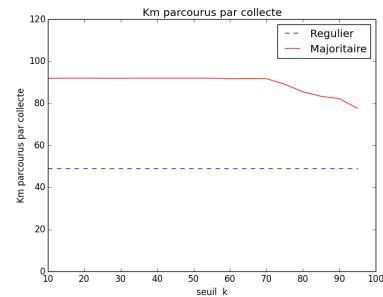


FIGURE 4 – Kilomètres parcourus en une collecte en fonction de k

Font partie du total des kilomètres :

- Les kilomètres du secteur pendant le ramassage
- Les allers retours en HLP pour vider la benne

Le palier dans le graphique en fonction de s est dû à une vidage de benne supplémentaire. On remarque que la benne fait toujours plus de kilomètres dans le modèle majoritaire que dans le modèle régulier. Ceci est expliqué par le fait que dans le modèle régulier, la benne ne ramasse qu'un secteur sur 4 à chaque collecte, ce qui fixe r à 3 : trois jours de remplissage, un jour de collecte. Par contre, dans le modèle majoritaire, jusqu'à $k \approx 70$, les 4 secteurs sont ramassés à chaque fois, la benne parcourt donc toutes les rues.

Cette simulation n'est pas très en faveur de la collecte à la demande. Cependant, en observant d'autres indicateurs tels que la quantité horaire de déchets récoltés, on observera un net avantage à cette collecte.

Par ailleurs, en regardant les différents graphiques qui décrivent les évolutions des indicateurs en fonction du seuil k , on constate à chaque fois un changement de comportement quand $k/n \approx 0.13$. Avant ce seuil, tous les secteurs sont systématiquement ramassés (ie prendre $k = 1$ ne changerait rien à ce moment),

tandis qu'après ce seuil, les secteurs n'ont pas à chaque fois assez de bacs pleins pour faire passer une benne. Nous avons essayé de comprendre ce phénomène, qui sont visibles quand le nombre de noeuds observés est grand (n grand).

5 Résultat de champ moyen

Ce résultat est resté au stade de conjecture.

Définition 5.0.1. Soit $N \geq 1$ le nombre de noeuds observés. Soit $(X^N(t))$ le processus majoritaire associé au variables de remplissage $(\alpha_{i,t})$, aux seuils de remplissage s et de ramassage k . On définit la mesure empirique d'occupation des noeuds de $(X^N(t))$ comme la mesure ν_t^N sur \mathbb{N} , définie pour $g : \mathbb{N} \mapsto \mathbb{R}+$ par :

$$\langle \nu_t^N, g \rangle = \frac{1}{N} \sum_{i=1}^N g(X^N(t)) \quad (8)$$

Notre conjecture concerne la convergence des ν_t^N en distribution, selon la valeur des k . On suppose que $(\nu^N(0))_N$ converge en distribution vers une mesure π_0 . On définit la chaîne de Markov $(Y(t))$ à valeurs dans \mathbb{N} , de distribution initiale π_0 qui correspond au processus majoritaire avec $n = 1$ par :

$$Y(t+1) = (Y(t) + \alpha_{1,t}) \mathbb{1}_{\{Y(t) < s\}} \quad (9)$$

On note π_t la loi de $Y(t)$.

Conjecture 5.0.1. On suppose que $k = \lceil \delta N \rceil$, avec $0 < \delta < 1$ qui vérifie :

$$\delta < \pi_0([s, \infty[) \quad (10)$$

alors $(\nu^N(t))_N$ converge en distribution vers π_t quand N tend vers l'infini. En particulier, presque sûrement, des noeuds sont remis à 0 à chaque instant.

Cette valeur critique de δ est tirée de la condition de ramassage. Il n'y a pas de ramassage au temps $t+1$ si l'événement suivant a lieu :

$$\mathcal{E}_t^N = \left\{ \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_i^N(t) \geq s\}} < \delta \right\} \quad (11)$$

6 Conclusion et ouvertures sur la suite

Ce stage en entreprise, également encadré par Philippe Robert, chercheur à l'INRIA, n'était pas très conventionnel. La réelle difficulté ne reposait pas forcément dans les résultats mathématiques, mais plutôt dans le processus de modélisation, qui a permis d'aboutir à un système mathématiques abordable et réaliste.

Les simulations nous ont notamment ouvert à de nouvelles pistes de réflexions, autant sur le plan mathématique que sur la précision du modèle à observer.

Malheureusement, ce stage limité dans le temps n'a pas permis de donner une réponse claire aux équipes de Vallée Sud- Grand Paris. Il a cependant ouvert de nombreuses réflexions sur les paramètres à prendre en compte si l'établissement souhaite poursuivre son idée de collecte à la demande.

Si une modélisation plus poussée était envisagée, voici quelques pistes de réflexions qui pourraient être abordées.

Tout d'abord, nous n'avons pas étudié le facteur géographique qui peut entrer en compte dans le trajet des bennes. Les trajets de collectes actuels sont optimisés pour ne passer qu'une seule fois dans chaque rue et parcourir tout les secteurs. Dans le cas d'une collecte à la demande, la benne devrait passer par des rues sans collecter pour se rendre d'une rue à collecter à une autre...

Ensuite, on pourrait s'intéresser aux impacts des diversités démographiques des villes : on peut imaginer que la collecte régulière est adaptée aux zones contenant une majorité d'immeubles, tandis que la collecte à la demande pourrait être avantageuse dans les zones pavillonnaires. Il faudrait déterminer à quel mesure ces suppositions sont vraies.

Références

- [1] Quentin Adrien Tonneau. *Optimisation de la chaîne logistique des déchets non dangereux*. Theses, Ecole nationale supérieure Mines-Télécom Atlantique, December 2017.