



HAL
open science

Transformation paramétrique de distance à vol d'oiseau en distance sur une voirie

Huong Nguyen, Frédérique Robin, Audrey Vonseel

► **To cite this version:**

Huong Nguyen, Frédérique Robin, Audrey Vonseel. Transformation paramétrique de distance à vol d'oiseau en distance sur une voirie. [Rapport de recherche] AMIES. 2018. hal-01949328

HAL Id: hal-01949328

<https://hal.science/hal-01949328>

Submitted on 10 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transformation paramétrique de distance à vol d’oiseau en distance sur une voirie

Huong Nguyen ^{*}, Frédérique Robin [†], Audrey Vonseel [‡]

August 21, 2018

1 Introduction

Ce rapport présente les résultats obtenus à l’issue de la semaine d’études maths-entreprise (SEME), organisée par l’Agence pour les Mathématiques en Interaction avec les Entreprises (AMIES) et le Laboratoire de Mathématiques et Applications (LMA) de l’Université de Poitiers et s’étant déroulée à Poitiers au printemps 2018.

L’étude, proposée par Catherine Gloaguen ¹ et encadrée par Hermine Biermé ², vise à proposer une estimation d’une distance Δ au plus court chemin sur une carte aléatoire modélisant des grands réseaux d’accès fixes sachant la distance à vol d’oiseau d . La modélisation aléatoire des rues utilise des techniques de géométrie stochastique (pavages de Poisson type Voronoï, Delaunay ou Ligne) et est considérée comme connue.

Après avoir rappelé brièvement le contexte et exposé les outils de géométrie stochastique exploités ainsi que la problématique, nous présenterons les trois démarches numériques proposées ainsi qu’une analyse de celles-ci. Des pistes pour une poursuite d’étude plus théoriques seront proposées en dernière partie.

Contents

1	Introduction	1
2	Contexte	2
2.1	Carte aléatoire	2
2.1.1	Processus ponctuel de Poisson	3
2.1.2	Pavage de Poisson-Voronoï	4
3	Démarche proposée	5
3.1	Précisions sur le pavage de Poisson-Voronoï	5
3.1.1	Domaine d’étude B	5
3.1.2	Intensité de saut λ_B	6
3.2	Distance à vol d’oiseau sur une carte aléatoire	7
3.2.1	Méthode 1 : Fixer un point sur un nœud	7
3.2.2	Méthode 2 : Tirage d’un point aléatoire	8
3.2.3	Méthode 3 : Tirage de points quelconques	8
4	Analyse des solutions proposées	9
4.1	Comparaison entre les trois méthodes	9
4.2	Proposition de modèles pour Δ en fonction de d	10
4.2.1	Modèle 1 : modèle linéaire	11
4.2.2	Modèle 2 : prise en compte de la taille d’une arête	12

^{*}Université Paris-Est, LAMA, UMR 8050, UPEMLV, CNRS, UPEC, F-77454 Marne-la-Vallée, France

[†]Inria, Université Paris-Saclay LMS, Ecole Polytechnique, CNRS, Université Paris-Saclay. mail: frederique.robin@inria.fr

[‡]Université de Strasbourg, CNRS, IRMA UMR 7501, F-67000 Strasbourg, France

¹Orange Labs, 38-40, rue du Général Leclerc, F-92794 Issy-les-Moulineaux, France

²Laboratoire de mathématiques et applications UMR CNRS 7348, 86962 Chasseneuil

5	Discussion/Généralisation	12
6	Annexe	13
6.1	Algorithme de Dijkstra	13
6.2	Effets de bords	14
6.3	Figure avec les nombres de nœuds traversés	14

2 Contexte

Des modèles stochastiques spatiaux pour des réseaux de télécommunications ont été développés depuis les dernières années comme une alternative à des modèles économiques traditionnels pour chiffrer les équipements et les plannings stratégiques [Gloaguen 2006]. Véritable structure vivante pouvant être considérée à différentes échelles : quartier, ville, région, etc., le réseau de télécommunication est une structure complexe d'une grande diversité de formes (circulaires, tentaculaires, droites, voir parfois même fractales [CT11]) non figées dans le temps. Soumis au grès de contraintes imposées par l'Homme, il s'auto-organise et se développe selon une logique de division et d'expansion.

En particulier, *Orange Labs* utilise les modèles de pavages aléatoires pour analyser le réseau routier dans le but d'optimiser les coûts de déploiements de réseaux de fibres sur le territoire [GC18]. Les ingénieurs ont souvent besoin d'estimer une distance Δ au plus court chemin en suivant la voirie entre deux points (nœuds du réseau, clients). On ne souhaite pas recourir à la distance sur une carte mais utiliser les modèles aléatoires à disposition. Par conséquent, ces points ne seront pas caractérisés par leurs positions géographiques mais par la distance entre ceux-ci. En exploitant les modèles aléatoires stationnaires simulant les rues, nous chercherons à proposer un modèle reliant les grandeurs Δ et d .

2.1 Carte aléatoire

Modéliser la géométrie des infrastructures (système routier) est la première étape dans l'analyse des réseaux de télécommunications. Il est donc important de choisir un modèle de pavage approprié [GFSS06]. Initialement proposé pour modéliser la répartition spatiale d'organismes vivants (par exemple, les capillaires sanguins [MEFS]), le processus de saut est également un outil mathématique pertinent pour rendre compte de la diversité des formes géométriques que l'on peut rencontrer dans un réseau routier. Lorsque l'on considère un intervalle de temps relativement court, le paysage urbain évolue peu et est dans un état quasi-stationnaire, ce qui motive l'utilisation de processus de saut homogène en temps.

La géométrie des infrastructures peut être rendu compte à l'aide d'un pavage du plan (partition sans superposition, ni trou), que l'on notera P . Ce pavage est supposé régulier, si bien que l'on peut le déterminer entièrement par trois grandeurs géométriques scalaires indépendantes, des intensités exprimées en unité d'aire, résumées dans le tableau ci-dessous :

intensité	interprétation	dimension
$\lambda^{(0)}$	Nombre moyen de noeuds p.u.a.	$[L]^{-2}$
$\lambda^{(1)}$	Nombre moyen de milieux de cotés p.u.a.	$[L]^{-2}$
$\lambda^{(2)}$	Nombre moyen de centres de cellules p.u.a.	$[L]^{-2}$
$\lambda^{(3)}$	Longueur moyenne totale des bords p.u.a.	$[L]^{-1}$

La morphologie d'un pavage est alors représentée par le vecteur

$$P = (\lambda^{(0)}, \lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}).$$

Bien que seul trois paramètres soient nécessaires pour caractériser un pavage, le dernier paramètre ($\lambda^{(3)}$, le nombre moyen d'arêtes) est inclus à des fins de vérification et doit être égale au nombre moyen de croisements plus les cellules.

Les modèles aléatoires de pavage sont des outils mathématiques privilégiés pour simuler efficacement un réseau à partir de règles simples et qui reproduisent la réalité (pour un nombre suffisant de réalisations). Parmi les processus existant, trois processus de pavage de Poisson sont généralement utilisés pour produire des pavages aléatoires dans le plan et dans l'espace : le pavage

de Poisson par des lignes (PLT : *Poisson Line Tessellation*), le pavage de Poisson-Voronoi (PVT : *Poisson Voronoi Tessellation*) et le pavage de Poisson-Delaunay (PDT : *Poisson Delaunay Tessellation*) [GFSS06], section 9.5 de [CSKM13]. Ces modèles Poissonniens, bien que simples, sont suffisants en général, mais de meilleurs ajustements peuvent être obtenus en considérant par exemple des pavages STIT (section 9.6 de [CSKM13]) ou en utilisant une procédure de segmentation qui partitionne automatiquement une ville en autant de parties homogènes que souhaité [GFSS06].

Ces trois processus sont spécifiés par une unique intensité λ que l'on peut relier aux paramètres caractérisant un pavage P par les relations ci-dessous :

P(type, λ)	PLT	PDT	PVT
$\lambda^{(0)}$ - Nombre moyen de nœuds p.u.a.	λ^2/π	λ	2λ
$\lambda^{(1)}$ - Nombre moyen de milieux de cotés p.u.a.	$2\lambda^2/\pi$	3λ	3λ
$\lambda^{(2)}$ - Nombre moyen de centres de cellules p.u.a.	λ^2/π	2λ	λ
$\lambda^{(3)}$ - Longueur moyenne totale des bords p.u.a.	λ	$32\sqrt{\lambda}/(3\pi)$	$2\sqrt{\lambda}$

Table 1: **Intensité des processus PVT, PLT et PDT et grandeurs caractéristiques.** Dans le cas de processus stationnaire, des relations algébriques peuvent être trouvées pour relier l'intensité de saut λ au grandeur caractéristique. Pour plus de détails, voir [VGS13] ainsi que le chapitre 10 de [SW08]

2.1.1 Processus ponctuel de Poisson

Dans cette partie, nous rappelons la définition d'un processus de Poisson dans \mathbb{R}^2 ainsi que quelques propriétés de celui-ci. Pour plus de détails, le lecteur peut se référer à la section 2.4 de [CSKM13].

Definition 2.1 (Processus ponctuel de Poisson dans \mathbb{R}^2 (cf [Fle03])) *On appelle X un processus ponctuel dans \mathbb{R}^2 d'intensité Λ_X si*

1. pour tout $n \geq 0$, pour tous B_1, B_2, \dots, B_n , ensembles mesurables bornés de \mathbb{R}^2 disjoints deux à deux, $(X(B_1), \dots, X(B_n))$ est un vecteur indépendant,
2. pour tout B mesurable de \mathbb{R}^2 , $X(B)$ suit une loi de Poisson de paramètre $\Lambda_X(B)$:

$$\mathbb{P}(X(B) = k) = e^{-\Lambda_X(B)} \frac{(\Lambda_X(B))^k}{k!}$$

Le point (1) de la définition 2.1 caractérise l'indépendance de la dispersion des points, tandis que le point (2) caractérise la distribution Poissonienne des points (section 2.4 de [CSKM13]).

Dans le cas d'un processus ponctuel stationnaire (tel que sont le PVT, PDT et PLT), la mesure $\Lambda_X(B)$ est proportionnelle à la mesure de Lebesgue μ : $\Lambda_X(B) = \lambda_X \mu(B)$. Le scalaire $\lambda_X \in (0, \infty)$ est l'intensité du processus. Cette intensité caractérise totalement le processus.

Remarque 2.1 (Processus de saut, autre caractérisation) *Un processus de saut peut être à la fois vu comme une suite de variables aléatoires $X = \{x_1, x_2, \dots\}$ ou une mesure de comptage : pour tout ensemble borélien C , $X(C)$ représente les points aléatoires de X dans C . (cf section 2.3.1 [CSKM13]).*

Remarque 2.2 [Stationnaire et isotrope]

1. Si $X = \{x_1, x_2, \dots\}$ est un processus de saut stationnaire, le processus translaté $X_x := \{x_1 + x, x_2 + x, \dots\}$, a la même distribution que X pour tout $x \in \mathbb{R}^2$ (cf section 2.3.1 [CSKM13]).
2. Un processus de saut $X = \{x_1, x_2, \dots\}$ est dit isotrope si, pour toutes rotations par rapport à l'origine θ , X et $X_\theta := \theta X = \{\theta x_1, \theta x_2, \dots\}$ ont la même distribution. Dans le cas d'une distribution stationnaire, cela revient à s'assurer que $\theta B := \{\theta b, b \in B\} = B$ (cf section 2.3.1 [CSKM13]).

Simulation numérique d'un processus de Poisson stationnaire dans \mathbb{R}^2 Un processus de Poisson stationnaire dans \mathbb{R}^2 se simule de la même manière qu'un processus dans \mathbb{R} . Pour plus de détails, voir section 2.5 de [CSKM13]. Pour cela, il suffit de se fixer une fenêtre d'observation $B \subset \mathbb{R}^2$. Par simplicité, B est choisi compact. On simule ensuite le nombre de points N dans le compact B selon une loi de Poisson de paramètre $\lambda\mu(B)$ (définition 2.1, point (2)). Puis on tire aléatoirement les N points selon une loi uniforme dans l'ensemble B (ces points sont indépendants les uns par rapport aux autres, définition 2.1, point (1)).

2.1.2 Pavage de Poisson-Voronoi

Parmi les trois pavages aléatoires élémentaires qu'étudie *Orange Labs*, nous avons choisi de nous concentrer sur le pavage de Poisson-Voronoi, qui est dédié généralement aux réseaux dans un quartier. Nous exposons ci-dessous comment simuler un tel processus.

On se fixe une intensité de saut λ_B et un ensemble $B \in \mathbb{R}^2$ supposé compact. Dans un premier temps, on simule un processus de Poisson stationnaire X dans \mathbb{R}^2 d'intensité $\lambda_B\mu(B)$. On note $X = \{x_1, x_2, \dots, x_N\}$ la suite de points générée par ce processus et N est le nombre de ces points aléatoires. Dans un second temps, on construit les sommets et les arêtes du pavage de Voronoï à partir de l'ensemble des centres X (cf Figure 1) : pour chaque centre $x_n \in X$, on construit sa cellule de Voronoï associée

$$p(x_n) := \bigcap_{m \neq n} H(x_n, x_m) = \{x \in \mathbb{R}^2 : |x - x_n| \leq |x - x_m|, \forall m \neq n\},$$

où H est le demi-plan défini par

$$x_n, x_m \in X, \quad H(x_n, x_m) := \{x \in \mathbb{R}^2 : |x - x_n| \leq |x - x_m|\}.$$

La mesure $P_{PVT}^{\lambda_B} := \{p(x_n) : x_n \in X(\omega)\}$ est le pavage de Poisson-Voronoi, construit à partir d'un processus de Poisson stationnaire d'intensité λ_B .

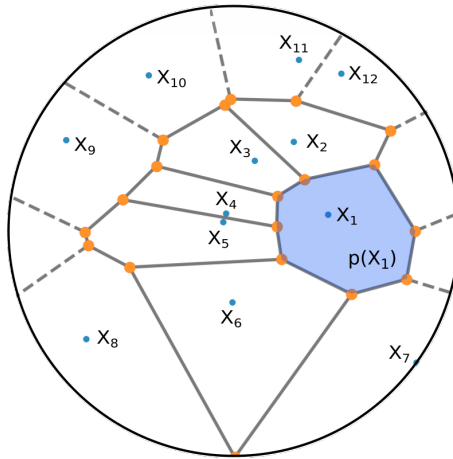


Figure 1: **Pavage de Poisson-Voronoi** : (Nombre de centres $N = 15$). Points bleus : centres x_n , réalisations d'un processus de Poisson dans un disque de centre 0 et de rayon 1; Région bleue : le polygone $p(x_1)$ centré en x_1 ; En orange : nœuds du graphe construits selon la procédure décrite ci-dessus.

Définition d'une carte aléatoire Nous définissons une carte aléatoire comme la construction d'une réalisation d'un processus de Poisson-Voronoi. Cette construction est un graphe du plan formé par les arêtes des cellules et les sommets d'une réalisation d'un pavage de Poisson-Voronoi P_{PVT} :

$$G(P_{PVT}^{\lambda_B}) := S \cup A, \quad S, A \subset P_{PVT}^{\lambda_B}$$

où S est l'ensemble des sommets et A l'ensemble des arêtes du pavage $P_{PVT}^{\lambda_B}$.

Distance sur une carte aléatoire Soient u et v deux points de P_{PVT} . Nous définissons les deux distances suivantes :

- la *distance réelle* entre u et v est la distance géodésique (longueur du plus court chemin) entre ces points dans le graphe, notée d_G (distance dans le graphe),
- la *distance à vol d'oiseau* entre u et v est la distance euclidienne entre ces points dans la carte, notée d_P (distance dans le pavage).

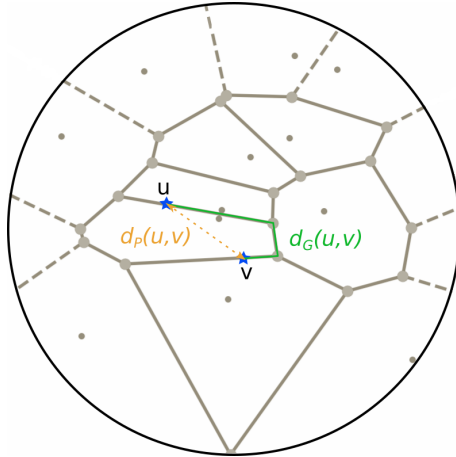


Figure 2: **Exemple de réalisation P** : En bleu : deux points u et v sur G . En vert : $d_G(u, v)$, la distance réelle entre u et v . En orange : $d_P(u, v)$, la distance à vol d'oiseau entre u et v .

3 Démarche proposée

Nous avons choisi une démarche exploratoire numérique afin d'étudier la relation existante entre la distance euclidienne d et la distance à vol d'oiseau Δ sur une carte aléatoire $P_{PVT}^{\lambda_B}$, caractérisée par une intensité λ_B (unité : p.u.a.) et un domaine B . Une approche naïve consiste en simuler un grand nombre de fois la carte $P_{PVT}^{\lambda_B}$ puis de calculer Δ et d et chercher à en déduire une relation entre ces deux variables aléatoires. Avant de présenter les solutions numériques retenues, nous présentons les simplifications pour notre démarche.

3.1 Précisions sur le pavage de Poisson-Voronoi

Afin de s'assurer de la cohérence des résultats et de leurs généralisations, une attention particulière doit être portée sur le choix des paramètres caractérisant le PVT : l'intensité λ (en p.u.a) du processus et le domaine B considéré. Nous décrivons ci-dessous les choix effectués.

3.1.1 Domaine d'étude B

La translation et le changement d'échelle (que nous décrirons ultérieurement) permettent de produire une séquence de points uniformément distribués dans tous rectangles ou hypercubes (p.39, [CSKM13]). Nous cherchons donc un domaine géométriquement simple à simuler, tel que la boule unité. Il ne reste plus qu'à définir le type : disque, carré, etc.

Pour s'affranchir de tous biais lié au domaine, nous proposons de choisir un processus isotrope. D'après la Remarque 2.2, un PVT possède cette propriété dès lors que B est invariant par rotation, et nous choisissons donc le disque unité $D(0, 1)$.

Transformation des coordonnées Dans [CSKM13], p.39, les auteurs proposent une manière de simuler des points aléatoires dans le disque unité $D(0, 1)$: un point aléatoire x tiré uniformément dans $D(0, 1)$ peut être décrit en coordonnées polaires :

$$x = (r, \theta), \quad r \in [0, 1] \text{ et } \theta \in]0, 2\pi].$$

Les variables aléatoires r et θ sont indépendantes et vérifient :

$$r = \sqrt{u_1} \text{ et } \theta = 2\pi u_2,$$

où u_1 et u_2 sont deux variables aléatoires uniformes sur $[0, 1]$.

3.1.2 Intensité de saut λ_B

Dans ce qui suit, nous raisonnons par rapport au domaine de référence $B = D(0, 1)$.

Ajustement du nombre de points L'intensité de saut λ_B dans un PVT correspond au nombre moyens de centres de cellules par unité d'aires (voir Tableau 1). Pour pouvoir construire un pavage de Voronoï, il est nécessaire qu'il y ait au minimum 3 centres, ce qui peut-être problématique pour des λ_B trop grands. En effet, si $\lambda_B \mu(B) = 1$, $\mathbb{P}[N \leq 2] = 0.99$. Pour augmenter le nombre moyen de centre, nous pouvons agrandir le domaine considéré.

Soit $\gamma_+ \in]1, +\infty[$ un coefficient correcteur. Si on considère l'ensemble $B = D(0, \gamma)$, alors l'intensité de saut du processus associé est $\lambda_{D(0, \gamma_+)} = \gamma_+ \pi > \lambda_{D(0, 1)}$. Le nombre moyen de centres de cellules dans l'ensemble $D(0, \gamma_+)$ est bien plus important que pour l'ensemble $D(0, 1)$.

A l'inverse, si $\lambda_{D(0, 1)}$ est trop grand, alors la carte aléatoire associée risque d'être très dense et d'avoir un grand nombre de connexions, ce qui ralentit le calcul des valeurs de Δ et d . La solution consiste alors en diminuer l'espace $D(0, 1)$ en multipliant l'intensité par $\gamma_- \in]0, 1[$.

Changement d'échelle Supposons $\lambda_{D(0, 1)}$ donnée. Si $d > 2$ (dans l'unité considéré), alors d n'appartient pas au domaine $D(0, 1)$: il est nécessaire d'effectuer un changement d'échelle ³.

Pour ce faire, nous construisons un PVT annexe, identique à celui de départ $P_{PVT}^{\lambda_{D(0, 1)}}$ mais exprimé dans une unité différente. Ce processus annexe $\tilde{P}_{PVT}^{\lambda_{D(0, 1)}}$ a donc la même intensité que celui de départ (donc, le même nombre de points) et le nouvel ensemble est une dilatation de celui de départ : $\gamma D(0, 1) := \{\gamma x, x \in D(0, 1)\}$ où γ est tel que

$$\Lambda_{D(0, 1)} = \Lambda_{\gamma D(0, 1)} \Rightarrow \lambda_{D(0, 1)} \pi = \lambda_{\gamma D(0, 1)} \gamma^2 \pi$$

On déduit donc $\lambda_{\gamma D(0, 1)} = \frac{\lambda_{D(0, 1)}}{\gamma^2}$.

Nombre de centres fixes Revenons à la manière dont un PVT est construit : 1) Simuler le nombre de centres N selon une loi de Poisson, 2) Tirer uniformément N points dans le disque unité $D(0, 1)$ et construire le pavage de Voronoï. On constate qu'il y a donc deux sources d'aléatoires indépendantes : la simulation du nombre de centre et la simulation de la position des points. S'il est compliqué d'étudier le positionnement des points, la loi du nombre de centre est totalement connue (définition 2.1) :

$$\mathbb{P}[N = k] = e^{-\lambda_{D(0, 1)} \pi} \frac{(\lambda_{D(0, 1)} \pi)^k}{k!}. \quad (1)$$

Par conséquent, chercher numériquement la loi du couple (Δ, d) sur $P_{PVT}^{\lambda_{D(0, 1)}}$, que nous notons $(\Delta, d)|P_{PVT}^{\lambda_{D(0, 1)}}$ peut se restreindre à chercher toutes les lois $(\Delta, d)|[P_{PVT}^{D(0, 1), n}, N = n]$. $P_{PVT}^{D(0, 1), n}$ désigne un PVT défini sur le domaine $B = D(0, 1)$ et dont le nombre de centre est fixé et égale à n . Nous proposons l'estimateur empirique de la loi du couple (Δ, d) ci-dessous :

$$(\hat{\Delta}, \hat{d})|P_{PVT}^{\lambda_{D(0, 1)}} = \sum_{n \in \mathcal{N}} \mathbb{E}[(\Delta, d)|P_{PVT}^{D(0, 1), n}, N = n] \times \mathbb{P}[N_{\lambda_{D(0, 1)}} = n], \quad (2)$$

sous l'hypothèse que $\mathbb{E}[(\Delta, d)|P_{PVT}^{D(0, 1), n}, N = n] < \infty$, pour tout $n \in \mathcal{N}$. L'ensemble \mathcal{N} est choisi de sorte que $\mathbb{P}[N_{\lambda_{D(0, 1)}} \in \mathcal{N}] = 0.95$.

En réalité, il serait plus intéressant pour les ingénieurs d'avoir l'estimateur de la loi marginale :

$$\hat{\Delta}|P_{PVT}^{\lambda_{D(0, 1)}}, d = \sum_{n \in \mathcal{N}} \mathbb{E}[\Delta|(P_{PVT}^{D(0, 1), n}, N = n, d)] \times \mathbb{P}[N_{\lambda_{D(0, 1)}} = n],$$

³Dans les faits, il vaut mieux faire un changement d'échelles dès que $d \geq 1$ afin de limiter les effets de bord.

Dans la partie suivante, nous présentons les algorithmes permettant d'estimer numériquement $\mathbb{E}[(\Delta, d)|P_{PVT}^{D(0,1),n}, N = n]$ ainsi que $\mathbb{E}[\Delta|(P_{PVT}^{D(0,1),n}, N = n, d)]$.

Pour simplifier les notations, nous noterons P^n le processus $P_{PVT}^{D(0,1),n}$ et G , le graphe issu de P^n ($G(P^n) = G$).

3.2 Distance à vol d'oiseau sur une carte aléatoire

Pour estimer numériquement $\mathbb{E}[(\Delta, d)|P_{PVT}^{D(0,1),n}, N = n]$ et/ou $\mathbb{E}[\Delta|(P_{PVT}^{D(0,1),n}, N = n, d)]$, nous devons échantillonner des points sur la carte aléatoire puis calculer la distance euclidienne d et la distance à vol d'oiseau Δ . Dans ce qui suit, la réalisation $P_{PVT}^{D(0,1),n}$ sera notée P^n pour plus de lisibilité. Nous proposons trois manières de déterminer ces quantités à travers différentes configurations.

3.2.1 Méthode 1 : Fixer un point sur un nœud

La première approche consiste à fixer la distance d et à relever Δ de la façon suivante :

Dans un graphe $G = S \cup A$, on fixe un sommet u_s dans S que l'on va appeler *point source*. Pour une distance d comprise dans $[0, 1]$ donnée, on détermine l'ensemble des points sur le graphe à distance à vol d'oiseau d du point source u_s . On détermine ensuite leur distance réelle au point source u_s à l'aide de l'algorithme de Dijkstra (pour une description de l'algorithme, voir section 6.1). Ce processus est répété pour une famille de points sources $\mathcal{U}_s^1 \subset S$.

1. Initialisation : un graphe $G = S \cup A$, une distance d
2. Construire une famille \mathcal{U}_s^1 de points sources dans S .
3. Pour tout sommet u_s de \mathcal{U}_s^1 :
 - (a) Trouver les points $u_e \in G$ tel que $d_P(u_s, u_e) = d$.
 - (b) Pour chaque u_e , calculer la distance réelle $d_G(u_s, u_e)$.

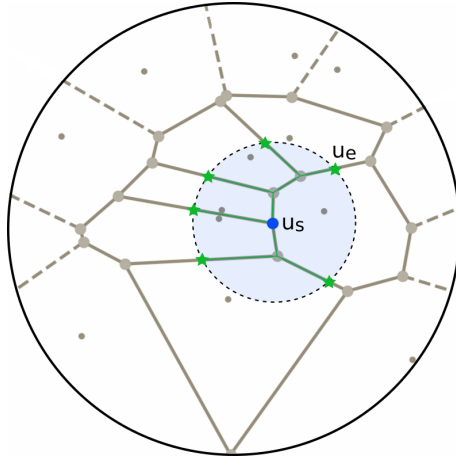


Figure 3: **Construction de points u_e à partir d'un point source-sommet u_s** : Point bleu : le point source u_s ; Région bleue : le disque de rayon d centré en u_s ; En vert : les points u_e de G à distance d de u_s .

Empiriquement, si l'on simule un certain nombre de fois P^n et que l'on effectue la démarche décrite ci-dessus, on obtient l'estimateur (empirique) pour $\mathbb{E}[\Delta|(P^n, N = n, d)]$ suivant :

$$\hat{\Delta}_1 = \hat{\mu}_1(d, P^n) = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{|\mathcal{U}_s^{1,i}|} \sum_{u_s \in \mathcal{U}_s^{1,i}} \left(\frac{1}{|\mathcal{C}(u_s, d)|} \sum_{u_e \in \mathcal{C}(u_s, d)} d_G(u_s, u_e) \right) \right), \quad (3)$$

où $|\mathcal{U}_s^{1,i}|$ est le cardinal de la famille $\mathcal{U}_s^{1,i}$ extraite d'une i -ème réalisation $P^{n,i}$ de P^n , $\mathcal{C}(x, r)$ est le cercle de centre x et de rayon r et $|\mathcal{C}(u_s, d)|$ est son cardinal.

3.2.2 Méthode 2 : Tirage d'un point aléatoire

La méthode 1 a l'inconvénient de fournir un sous-échantillonnage dû à la restriction du choix du point source aux sommets du graphe. La seconde approche que nous avons étudiée reprend la construction de la méthode 1 en choisissant de façon aléatoire la position du point source dans le graphe G .

1. Initialisation : un graphe G , une distance d
2. Construire une famille \mathcal{U}_s^2 de points sources dans G .
3. Pour tout point u_s de \mathcal{U}_s^2 :
 - (a) Trouver les points $u_e \in G$ tel que $d_P(u_s, u_e) = d$.
 - (b) Pour chaque u_e , calculer la distance réelle $d_G(u_s, u_e)$.

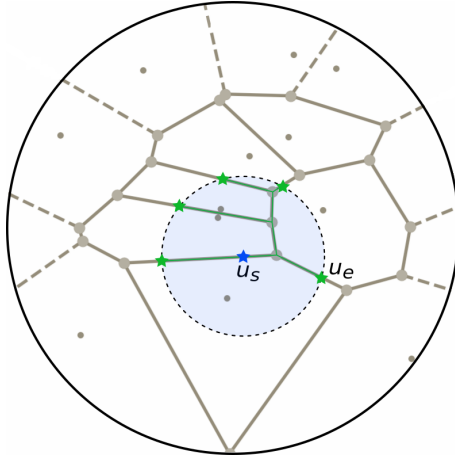


Figure 4: **Construction de points u_e à partir d'un point source u_s sur une arête** : Point bleu : le point source u_s ; Région bleue : le disque de rayon d centré en u_s ; En vert : les points u_e de G à distance d de u_s .

De la même manière que pour la méthode 1, nous proposons un autre estimateur empirique pour $\mathbb{E}[\Delta|(P^n, N = n, d)]$:

$$\hat{\Delta}_2 = \hat{\mu}_2(d, P^n) = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{|\mathcal{U}_s^{2,i}|} \sum_{u_s \in \mathcal{U}_s^{2,i}} \left(\frac{1}{|\mathcal{C}(u_s, d)|} \sum_{u_e \in \mathcal{C}(u_s, d)} d_G(u_s, u_e) \right) \right), \quad (4)$$

où $|\mathcal{U}_s^{2,i}|$ est le cardinal de la famille $\mathcal{U}_s^{2,i}$ extraite d'une i -ème réalisation $P^{n,i}$ de P^n , $\mathcal{C}(x, r)$ est le cercle de centre x et de rayon r et $|\mathcal{C}(u_s, d)|$ est son cardinal.

3.2.3 Méthode 3 : Tirage de points quelconques

Cette troisième approche consiste à observer les distances réelle et à vol d'oiseau de couple de points quelconque sur une carte. A l'inverse des deux précédentes méthodes, nous pouvons estimer la loi du couple (i.e. estimer $\mathbb{E}[(\Delta, d)|P_{PVT}^{D(0,1),n}, N = n]$) au lieu de la loi jointe.

Comme dans les méthodes précédentes, on suppose que l'on dispose d'un graphe G construit à partir d'une réalisation de P^n . On choisit ensuite de façon uniforme un ensemble de points sur cette carte. Pour tous les couples (u_i, v_i) , $i \in \mathcal{I}$, de points de cet ensemble, on détermine la distance dans le graphe $\Delta_i = d_P(u_i, v_i)$ et la distance euclidienne $d_i = d_G(u_i, v_i)$.

1. Initialisation : un graphe G , une famille \mathcal{U}^3 de couples de points de G
2. Pour tout couple de points u, v de \mathcal{U}^3 , calculer :
 - la distance réelle $d_G(u, v) = d$,

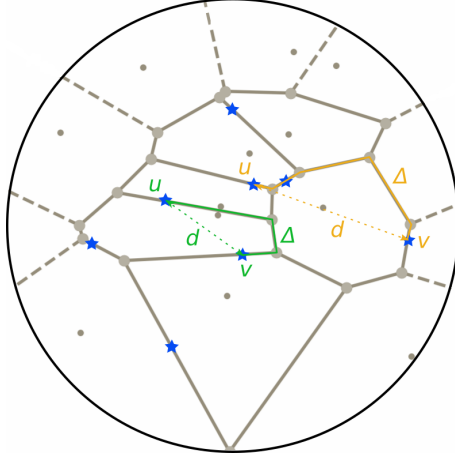


Figure 5: **Exemple de relevé de d_i et Δ_i pour des couples de points de \mathcal{F}** : En bleu : les points de \mathcal{F} - En vert et orange : deux exemples de relevé de d et Δ .

- la distance à vol d'oiseau $d_P(u, v)$.

Naïvement, un estimateur empirique que l'on peut proposer pour $\mathbb{E}[(\Delta, d)|P^n, N = n]$ est :

$$(\hat{\Delta}_3, \hat{d}_3) = \hat{\mu}_3(P^n) = \left(\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{|\mathcal{U}^{3,i}|} \sum_{(u,v) \in \mathcal{U}^{3,i}} d_G(u, v) \right), \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{|\mathcal{U}^{3,i}|} \sum_{(u,v) \in \mathcal{U}^{3,i}} d_P(u, v) \right) \right), \quad (5)$$

où $|\mathcal{U}^{3,i}|$ est le cardinal de la famille $\mathcal{U}^{3,i}$ extraite d'une i -ème réalisation $P^{n,i}$ de P^n .

L'estimateur ci-dessus ne propose qu'une valeur moyenne du couple (Δ, d) , ce qui ne correspond pas réellement aux attentes des ingénieurs d'*Orange Labs*. Nous verrons dans la prochaine section des pistes pour construire un estimateur de Δ_3 à partir de d avec cet algorithme.

4 Analyse des solutions proposées

Les trois méthodes présentées ci-dessus ont chacune leurs avantages et leurs inconvénients. Dans cette section, nous analysons ces éléments et proposons deux modèles pour estimer Δ en fonction de d .

4.1 Comparaison entre les trois méthodes

Même pour un nombre de centre N identiques, deux réalisations d'un PVT P^N peuvent être visuellement très différentes, comme le montre la Figure 4.1. Cette différence est également visible lorsque l'on regarde les distributions des tailles d'arêtes (Figure 10). Pour pouvoir comparer nos trois méthodes de calculs de distance moyenne Δ en fonction de d , il est donc nécessaire de se placer à environnement fixe, i.e. comparer les résultats pour chaque graphe G_i construit à partir d'une réalisation d'un PVT à nombres de cellules n fixé, $P^{n,i}$.

Génération des jeu des données Nous avons généré 10 graphes : 5 avec un nombre de centre $N = 15$ et 5 autres avec un nombre de centre $N = 30$ selon la méthode exposée dans la partie (3.1.2). Pour chacun de ces graphes, nous avons calculé les distances Δ et d pour chacune des méthodes selon les algorithmes présentés en partie (4.2). Pour la méthode 1, la famille \mathcal{U}_s^1 est choisi comme l'ensemble total des sommets du graphe considéré. Pour la méthode 2, nous avons tiré 200 points pour un d fixé dans le graphe. Pour les méthodes 1 et 2, nous avons regardé 101 d différents répartis uniformément sur $]0, 2]$. Pour finir, 10 500 points (soit 10 500 * (10 500 - 1) couples) ont été tiré par graphe pour la méthode 3. Les résultats par nombre de centre N sont exposés dans la Figure 4.2.

Dans ce qui suit, nous comparons les distributions obtenues avec les méthodes 1, 2 et 3, que nous notons respectivement $\Delta_1^n|d$, $\Delta_2^n|d$ et (Δ_3^n, d_3^n) .

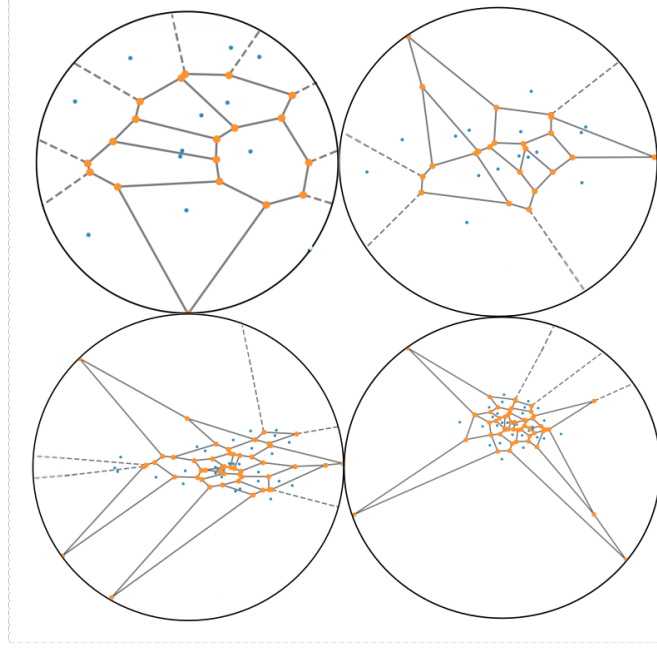


Figure 6: **Réalisations de P^N** . Illustrations de la variabilité des cartes obtenues intra-groupe (même nombre de centres N) et inter-groupe (entre deux groupes de N différents). En haut : $N = 15$, en bas : $N = 30$.

Méthode 1 vs Méthode 2 Pour ces deux méthodes, d n'est pas aléatoire. Par conséquent, nous utilisons un test de rang (Mann-Whitney) pour établir si les distributions $\Delta_1|d$ et $\Delta_2|d$ sont identiques ou non pour chacune des valeurs de d considérées. Bien que les résultats soient variables selon les valeurs de d , on peut considérer que les distributions $\Delta_1|d$ et $\Delta_2|d$ sont différentes. Ce résultat n'est pas surprenant puisque $\Delta_1|d$ est construit à partir d'un sous-échantillon de points, en comparaison de $\Delta_2|d$.

Méthode 2 vs Méthode 3 La comparaison entre la Méthode 2 et la Méthode 3 est délicate du faite que l'une calculée pour des valeurs ponctuelles de d alors que cette valeur varie avec la Méthode 3. Pour pouvoir comparer les distributions, nous proposons de clusteriser les couples (Δ_3, d_3) tel que :

$$\Delta_3|[d - \epsilon, d + \epsilon] := \{\Delta_3, d_3 \in [d - \epsilon, d + \epsilon]\},$$

où ϵ est choisi aussi petit que possible. Ce faisant, nous pouvons à présent effectuer un test de rang entre les distributions $\Delta_2|d$ et $\Delta_3|[d - \epsilon, d + \epsilon]$, similairement au paragraphe ci-dessus. Pour $\epsilon = 0.005$, les distributions $\Delta_2|d$ et $\Delta_3|[d - \epsilon, d + \epsilon]$ sont globalement différentes. Une hypothèse pour expliquer cet écart serait que les distributions n'ont pas été suffisamment échantillonnées, mais il est également possible que ce ne soit pas la seule raison.

Pour conclure cette sous-section, nous retiendrons que les méthodes 1 et 2 fournissent des résultats différents, ce qui est attendu. Les méthodes 2 et 3 semblent également différentes mais il serait nécessaire de confirmer ce résultat.

4.2 Proposition de modèles pour Δ en fonction de d

Bien que nos trois méthodes produisent des distributions statistiquement différentes, nous pouvons tout de même constater un comportement qualitatif semblable : Δ est globalement proportionnel à d . Nous pouvons même aller plus loin et constater qu'une bi-modalité est observée : pour des petites valeurs de d (entre 0 et 0.2), Δ est très proche de d tandis que lorsque d augmente, Δ est strictement supérieur à d .

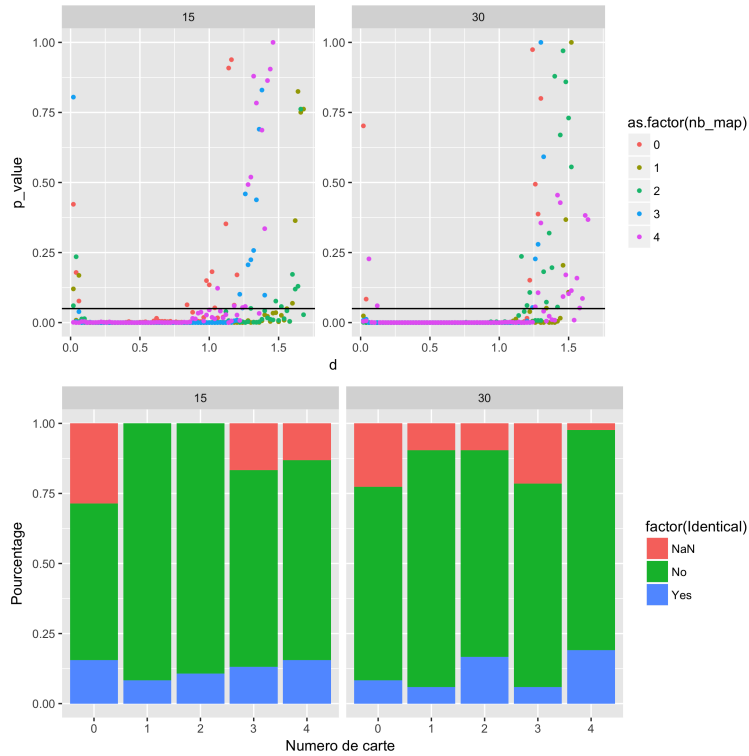


Figure 7: **Comparaison entre les distributions $\Delta_1|d$ et $\Delta_2|d$ (test de Mann-Whitney).** Pour chaque valeur de d uniformément répartie dans $]0, 2]$, un test de rang a été effectué entre les distributions $\Delta_1|d$ et $\Delta_2|d$. *Haut* : p-value du test en fonction de d par nombre de centre $N = 15, 30$ et réalisations P^N . La ligne noire correspond à un seuil de 95%. Si l’hypothèse nulle (H_0 : distributions identiques) est acceptée dans quelques cas pour des d petit (< 0.04) et grand (> 0.8), la tendance générale est au rejet de l’hypothèse nulle (hypothèse alternative : les distributions sont différentes). *Bas* : Pourcentage de distributions identiques en fonction du numéro de carte selon le nombre de centres N . Globalement, les distributions $\Delta_1|d$ et $\Delta_2|d$ ne sont pas identiques. Les résultats de type *NaN* indiquent les cas où pour une valeur d donnée, la distribution $\Delta_1|d$ ou $\Delta_2|d$ n’a pas pu être observée. Ce cas de figure peut arriver pour des valeurs de d trop grandes ($d > 1.5$).

4.2.1 Modèle 1 : modèle linéaire

En première approximation, nous proposons de modéliser par une fonction linéaire le lien entre Δ et d . Pour chaque valeur N , nous avons estimé le facteur correctif α_N de la relation

$$\Delta = \alpha_N d$$

à l’aide d’une régression linéaire effectuée avec le logiciel R. Les résultats sont proposés dans la Tableau 4.2.1. Un test de Student effectué sur les paramètres montre que α_N dépend de N .

	N = 15		N = 30	
Méthode	α	Erreur standard	α	Erreur standard
1	1.21	0.0009	1.21	0.0004
2	1.28	0.0002	1.25	0.0001
3	1.32	0.0009	1.27	0.0006

Table 2: **Estimation du facteur correctif α pour $N = 15$ et $N = 30$.** Un test de comparaison de régression linéaire (test de Student sur les coefficients directeurs) indique que les coefficients α sont significativement différents entre eux. Par conséquent, d’une part, les trois méthodes sont différentes et d’autre part, le coefficient α dépend du nombre de centre N .

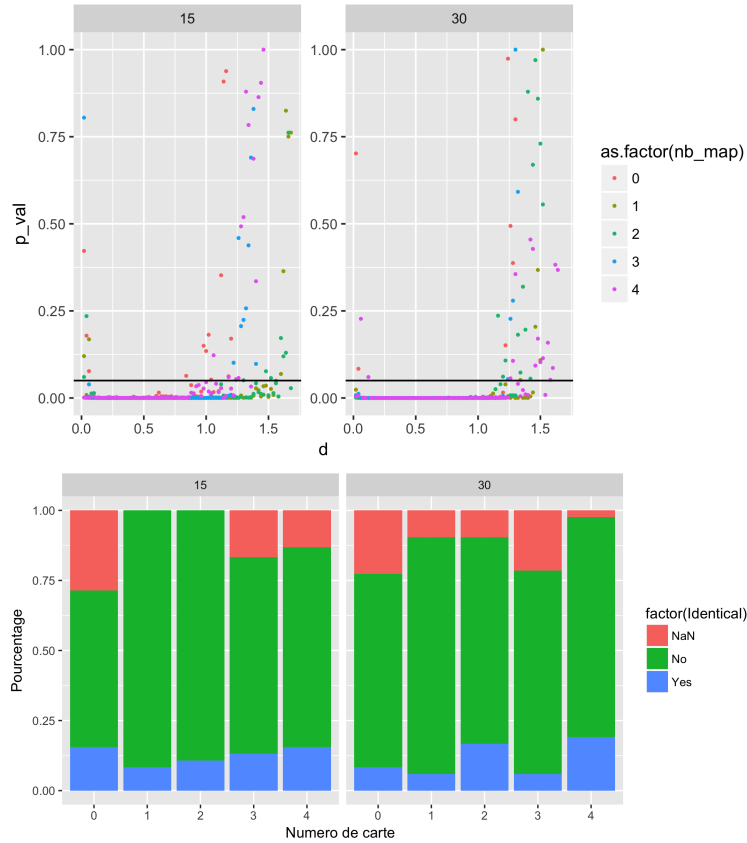


Figure 8: **Comparaison entre les distributions $\Delta_2|d$ et $\Delta_3|[d - \epsilon, d + \epsilon]$ (test de Mann-Whitney).** Pour chaque valeur de d uniformément répartie dans $]0, 2]$, un test de rang a été effectué entre les distributions $\Delta_2|d$ et $\Delta_3|[d - \epsilon, d + \epsilon]$ ($\epsilon = 0.005$). La construction ainsi que les conclusions de cette figure sont identiques à la Figure 4.1

4.2.2 Modèle 2 : prise en compte de la taille d'une arête

L'écart entre d et Δ en fonction de N dépend de manière évidente de l'écart entre d et la taille d'une arête. La Figure 6.3 nous indique que le nombre de noeuds traversés pour calculer la distance au plus courts chemin Δ augmente avec la distance d , ce qui rend-compte de l'écart entre d et Δ . En ajoutant la variabilité des distributions des tailles d'arêtes observée pour des graphes de même N ou non (Figure 10), l'hypothèse que cet écart est lié avec la taille des arêtes des cartes observées semble plus que raisonnable.

Soit n , un nombre de centre fixé. Nous définissons $L|N = n$ comme la loi d'une taille d'arête de P^n . Nous pouvons alors directement faire le lien entre Δ^n et d^n en fonction de cette loi :

$$\Delta^n = d^n \text{ avec probabilité } \mathbb{P}[L|N = n \geq d^n].$$

On a donc bien pris en compte le phénomène suivant : lorsque d augmente, la probabilité que d soit de taille inférieur ou égale à une arête du graphe diminue.

Le problème réside à présent en calculer $\mathbb{P}[L|N = n \geq d]$. Il serait envisageable d'estimer numériquement la loi de $L|N = n$ en simulant un grand nombre de fois P^n et en calculant la tailles des arêtes. Pour finir, il est envisageable d'utiliser les changements d'échelle vu en partie 4.1.2 pour ajuster les valeurs de d en fonction de $\mathbb{P}[L|N = n \geq d]$.

5 Discussion/Généralisation

Dans ce rapport, nous considérons trois approches numériques pour explorer des pistes de modélisations reliant la distance à vol d'oiseau Δ à la distance euclidienne d sur une carte aléatoire. Nous avons proposé trois estimateurs empiriques (3), (4) et (5). La qualité de ces estimateurs (sans biais

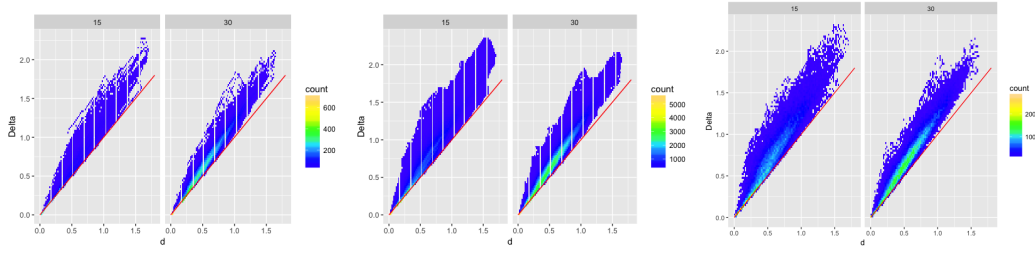


Figure 9: d en fonction de Δ pour différentes valeurs de n

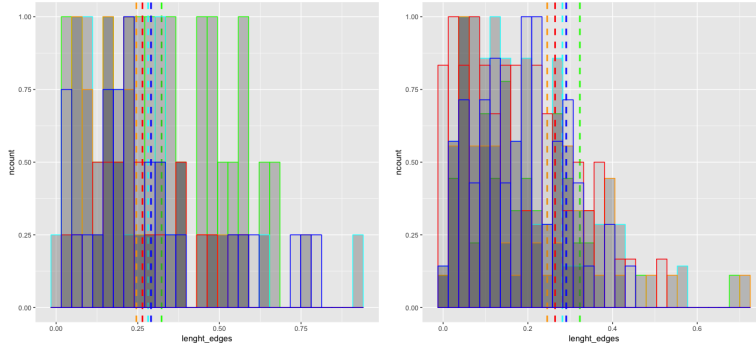


Figure 10: Taille moyenne d'une arête par carte et pour différentes valeurs de n

et convergent) n'a pas été exploré au cours de la semaine.

Bien que ces trois méthodes fournissent des distributions $\Delta|d$ différentes à environnement fixé, elles s'accordent sur l'existence d'une relation de proportionnalité entre d et Δ . En tenant compte de la loi d'une arête $L|N = n$ sur une réalisation d'un PVT de n cellules, P^n , et en jouant sur les changements d'échelles, nous pensons qu'il est possible de construire un estimateur reliant $\Delta|d$ plus pertinent que la relation linéaire que nous avons proposé en première approximation. Il sera alors utile de s'intéresser aux effets de bords (partie 6.2), que nous n'avons pas considéré dans notre analyse préliminaire du jeu de données.

Cette démarche peut par ailleurs s'étendre à d'autres processus de Poisson stationnaire dans le plan, tel que le PLT ou le PDT. En particulier, pour le PDT, la fonction de Ripley pourrait aider à déterminer la loi de $L|N = n$ [MEFS].

6 Annexe

6.1 Algorithme de Dijkstra

L'algorithme de Dijkstra [Dij59] répond au *problème du plus court chemin*, c'est-à-dire : étant donné un graphe (orienté ou non) et deux sommets de ce graphe, trouver un chemin entre les deux sommets dans le graphe, de poids minimum.

Soit \mathcal{G} un graphe. Soient A et B deux sommets de \mathcal{G} .

Principe : Le plus court chemin joignant A à B passe par un plus court chemin joignant A à chacun des sommets traversés. On peut donc construire de proche en proche un chemin le plus court joignant A à B . A chaque itération de l'algorithme, on choisit parmi les sommets non testés celui qui est à distance minimale du dernier sommet choisi.

Algorithme :

Initialisation : La distance du sommet de départ à lui-même est 0. On affecte la valeur ∞ à la distance à A provisoire aux autres sommets. On marque l'ensemble des sommets comme non visités. Le sommet de départ A est le sommet courant.

Répéter les opérations suivantes :

1. Considérer l'ensemble des sommets adjacents au sommet courant et calculer leur distance à A provisoire en traversant le sommet courant. A chacun de ces sommets, assigner la valeur minimal entre la distance à A provisoire calculée et celle précédemment attribuée. Marquer le sommet courant comme visité.
2. Si le sommet d'arrivé B a été marqué visité ou si la plus petite distance à A provisoire des sommets non visités est infinie (il n'existe pas de chemin joignant A à B), l'algorithme s'arrête.
3. Sinon, sélectionner le sommet non visité de distance à A provisoire minimale comme sommet courant.

6.2 Effets de bords

Le choix des ensembles \mathcal{U}_s^1 et \mathcal{U}_s^2 doit tenir compte d'un risque de sous-échantillonnage possible lorsque le nœud source u_S est tiré trop proche du bord. Les algorithmes des méthodes 1 et 2 n'explorent alors pas tout l'espace disponible, c'est-à-dire le cercle $\mathcal{C}(u_S, d)$, mais seulement une proportion $\mathcal{K}_p(u_S, d)$ de cet espace. Selon la valeur de d , cet effet peut-être plus ou moins important (Figure 6.2). Lorsque $d > 0.75$, ce qui apparaît comme une distance relativement faible (75% du rayon), environ 75% de l'échantillon a subit un effet de bord.

Pour exploiter les données calculées, on peut envisager une pondération en fonction du pourcentage de surface disponible :

$$\frac{\int_{\mathcal{K}_p(u_S, d)} dx}{\int_{\mathcal{C}(u_S, d)} dx}.$$

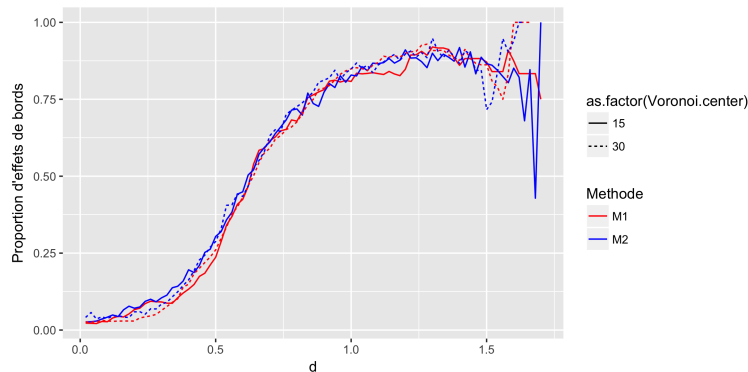


Figure 11: **Étude des effets de bords** Proportion d'effets de bords observées dans le jeu de données en fonction de d selon le nombre de centre N de cellules de Voronoï et la méthode considérée (méthode 1 ou méthode 2). L'effet de bord ne semble pas dépendre ni du nombre de centre N , ni de la méthode choisie mais totalement de la distance d .

6.3 Figure avec les nombres de nœuds traversés

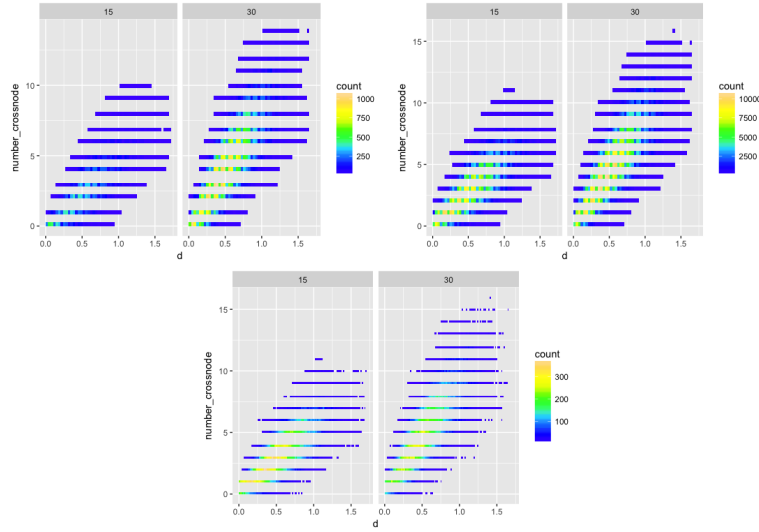


Figure 12: **Nombre de nœuds traversé en fonction de la distance d .** Pour chaque calcul de distance $d_P(u, v)$, nous comptons le nombre de nœuds du réseau traversés. Par exemple, si $d_P(u, v) = d_P(u, n) + d_P(n, v)$, où $n \in G$ un nœud tel que $(u, n) \in A$ (resp. $(n, v) \in A$), un unique nœud sera comptabilisé.

References

- [CSKM13] Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 3rd edition, 2013.
- [CT11] Gloaguen C. Courtat T., Douady S. Mathematics and morphogenesis of cities: a geometrical approach. *Phys Rev E* 83(3):15:26, 2011.
- [Dij59] Edsger W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [Fle03] Frank Fleischer. *Analysis and fitting of random tessellation models: applications in telecommunication and cell biology*. PhD Thesis, Fakultät für Mathematik und Wirtschaftswissenschaften, 2003.
- [GC18] Catherine Gloaguen and Elie Cali. Cost estimation of a fixed network deployment over an urban territory. *Annals of Telecommunications*, 73(5-6):367–380, 2018.
- [GFSS06] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. Fitting of stochastic telecommunication network models via distance measures and Monte–Carlo tests. *Telecommunication Systems*, 31(4):353–377, April 2006.
- [MEFS] T. Mattfeldt, S. Eckel, F. Fleischer, and V. Schmidt. Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. *Journal of Microscopy*, 223(2):107–119.
- [SW08] Schneider Rolf and Wolfgang Weil. *Stochastic and Integral Geometry*. Springer edition, 2008.
- [VGS13] Florian Voss, Catherine Gloaguen, and Volker Schmidt. Random Tessellations and Cox Processes. In Evgeny Spodarev, editor, *Stochastic Geometry, Spatial Statistics and Random Fields*, volume 2068, pages 151–182. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.