

GÉOMATHÉMATIQUE DES FLUX : INFLUENCE DU MAILLAGE SUR LES STATISTIQUES DE DÉPLACEMENTS

Christophe TERRIER (*)

(*) Insee, Département de la coordination statistique

Introduction

Toutes les enquêtes réalisées pour mesurer les déplacements des personnes se réfèrent -pas toujours de façon explicite- à des objets géographiques. Par exemple on pourra mesurer un flux de véhicules automobiles en se référant à cet objet linéaire qu'est la route. Pour les flux de personnes, la plupart des enquêtes de déplacements dites « ménages » -c'est à dire menées auprès d'individus ou de groupes d'individus-, prennent comme référence des territoires - entendus ici comme des portions délimitées de l'espace - entre lesquels on se déplace. Les règles de calcul applicables ne seront pas les mêmes selon le type d'objet géographique auquel on se réfère, d'où l'expression employée ici de *géomathématique des flux*. On en décrira les principes pour chaque type d'objet géographique et les conséquences sur les résultats obtenus.

La partie originale présentée ici porte sur l'étude de l'influence du maillage sur les distances mesurées dans les enquêtes sur les déplacements. Ce papier reprend, pour la bonne compréhension du cadre dans lequel on opère, quelques notions de base déjà décrites dans des papiers précédents¹.

1. Les objets géographiques utilisés pour la mesure des flux

Les objets géographiques utilisés pour la mesure des flux sont soit des lignes (voies de circulation) soit des territoires. Le point joue également un rôle, soit en tant que tel (point de passage, par exemple), soit comme limite sur une ligne (point de départ, point d'arrivée), soit employé par approximation à la place d'un territoire, en particulier lorsqu'on assimile ce territoire à son centre ou à son chef-lieu.

La circulation s'effectuant sur des axes de circulation (routes, voies ferrées, couloirs aériens), l'objet géographique naturel pour la mesure des flux devrait être la ligne tandis que le territoire serait dévolu à la mesure de l'afflux, c'est-à-dire de la présence de personnes à un moment donné sur ce territoire. En pratique, l'essentiel -voire la totalité- des statistiques obtenues par des enquêtes sociodémographiques se réfèrent à des territoires.

1.1. Les autres objets : lignes et points

Ces autres objets ne sont évoqués ici que pour établir le cadre général mais ils ne donneront pas lieu à des développements spécifiques dans ce papier. L'essentiel du texte suivant est repris de l'article déjà cité.

1.1.1. Les mesures basées sur des lignes

¹ [9] Flux et afflux de touristes : les instruments de mesure, la géomathématique des flux [Terrier 2006]

Les lignes qui nous intéressent ici sont les voies de communication ; la mesure des flux consiste à dénombrer les éléments qui circulent sur ces voies –c'est à dire les véhicules. Or, ceux-ci sont dotés d'une autonomie plus ou moins grande selon le type de voie et selon le mode de transport. On est donc amené à distinguer des lignes « fermées » et des lignes « ouvertes ».

Une ligne sera dite fermée lorsqu'un élément qui y circule ne dispose pas d'autonomie dans son trajet. La ligne de chemin de fer en est la forme la plus pure : elle est matérialisée par un double ruban d'acier que les véhicules ne peuvent pas quitter. Les trajets s'effectuent de façon très prévisible le long de la voie ferrée, depuis le point de départ jusqu'au point d'arrivée. Mais la ligne n'est strictement fermée que d'un point de départ unique à un point d'arrivée unique ; une correspondance avec changement de train au cours du trajet implique donc une rupture de ligne et un train omnibus, bien qu'il ne quitte jamais la voie ferrée, doit être considéré comme empruntant non pas une ligne fermée mais une succession de lignes mises bout à bout.

Les lignes aériennes sont, en pratique, définies par leur point de départ et leur point d'arrivée. La connaissance de la trajectoire réelle de l'avion et des couloirs aériens empruntés ne présente pas d'intérêt réel, en tout cas pour la mesure des flux. On peut donc considérer les lignes aériennes comme fermées.

La route représente, elle, un cas de ligne très ouverte. D'une part, le réseau en est d'une grande complexité, permettant d'emprunter de multiples chemins pour aller d'un point à un autre. D'autre part, les véhicules qui l'empruntent disposent d'une grande autonomie et sont en capacité à tout moment de bifurquer, de s'arrêter ou de faire demi-tour. Il n'y a donc pas de relation évidente entre les lignes et les trajets. L'autoroute peut cependant, sur certains tronçons entre péages ou accès, canaliser strictement la circulation, se rapprochant ainsi du concept de ligne fermée.

Seule une ligne fermée se prête véritablement à la mesure des flux. D'une part, les déplacements effectués sur des lignes fermées font l'objet d'une gestion et d'un enregistrement –en général par billetterie - qui rendent inutiles d'autres types de dénombrement. D'autre part, l'information est la même qu'on la mesure à l'origine, à destination ou en un point quelconque du parcours. Si l'on souhaite dépasser la simple mesure des flux et procéder à une enquête auprès des passagers, on peut donc la mener au lieu le plus adapté, différent selon le mode de transport : pour l'avion, on enquête généralement dans la salle d'embarquement ; pour le chemin de fer, les enquêtes se font plutôt dans le train au cours du voyage car il n'y a pas d'attente avant le départ.

Sur les lignes ouvertes, les possibilités de mesure sont réduites. Ces lignes se caractérisent par une liberté de mouvement –entrée, sortie, arrêt, bifurcation, demi-tour - des véhicules qui les empruntent. On sait mesurer les passages en un point donné, par des compteurs automatiques, des péages ou des dénombrements réalisés par des enquêteurs, mais un flux mesuré en un point du trajet ne permet pas de déduire directement le flux en un autre point. En pratique, on découpe les routes en tronçons sur lesquels la circulation est à peu près homogène. On affecte à l'ensemble du tronçon les flux mesurés en un point. Les zones urbaines, trop ouvertes, restent en dehors de ce système.

Il n'y a pas véritablement de calculs qui puissent s'opérer sur des lignes au-delà de la simple addition des flux mesurés successivement sur un même tronçon. On peut cependant chercher à extrapoler, à une ligne ou à un territoire, les flux mesurés en un point. Comme toute extrapolation, cette opération nécessite de faire des hypothèses de comportement, si possible basées sur des informations complémentaires dont on dispose par ailleurs. Comme noté précédemment à propos de ce genre de démarche, la fiabilité des résultats dépend de celle des informations complémentaires utilisées pour construire des hypothèses plausibles de comportement.

Dans la pratique, ce sont les autoroutes qui posent le plus de problèmes pour les enquêtes. Par exemple, intéressons-nous aux flux des touristes étrangers circulant en voiture en France. On fera l'hypothèse préalable -mais discutable- qu'il y a identité entre le pays de résidence des personnes et le pays d'immatriculation du véhicule. Il est possible de compter, par nationalité d'immatriculation, le nombre de véhicules franchissant tel péage sur telle période. Sur la base de dénombrements effectués par des enquêteurs, on pourra extrapoler ces comptages de véhicules pour estimer des flux de voyageurs. Laissons de côté les difficultés annexes de ces dénombrements, causées par le développement des vitres teintées et l'uniformisation grandissante des immatriculations dans les pays d'Europe. Remarquons au passage que les péages présentent deux avantages : - les véhicules s'y arrêtent, ce qui permet à un enquêteur de noter quelques caractéristiques comme le pays d'immatriculation et le nombre de passagers - les conducteurs acquittent un droit de passage pour lequel un certain nombre d'entre eux utilisent un moyen de paiement automatique qui permet d'identifier leur pays d'origine. Laissons encore de côté les difficultés causées par les télépéages - les véhicules ne s'arrêtent plus mais ils ralentissent quand même- et par les fortes disparités de

comportement selon la nationalité dans l'usage de la carte bancaire. A l'aide de ces dénombrements on sera assez bien informé sur les flux de véhicules empruntant tel ou tel tronçon d'autoroutes mais, compte-tenu de la multiplicité des péages, des entrées et des sorties, on ne saura pas répondre à quelques questions simples : Combien d'Allemands ont traversé la France d'une traite sans s'arrêter notablement avant la frontière espagnole ? Combien ont effectué des séjours d'étape ? Combien ont séjourné dans telle ou telle région ?

Si maintenant on souhaite en savoir un peu plus sur ces visiteurs de l'étranger, il va falloir attendre qu'ils s'arrêtent eux-mêmes et là, ne restent plus que les enquêtes sur les aires de repos avec tous les problèmes de représentativité que l'on peut imaginer².

1.1.2. Les mesures basées sur des points

Les points ne sont cités ici que pour mémoire. Dans la réalité, lorsqu'il effectue un déplacement, un individu va d'un point de départ à un point d'arrivée en suivant un chemin, c'est à dire une ligne. Dans la pratique des enquêtes actuelles on ne relève pas exactement les coordonnées des points de départ et d'arrivée mais on se contente de connaître les territoires respectifs auxquels appartiennent ces points et c'est donc le territoire qui sert de référence à la mesure des déplacements. De même on ne relève pas le cheminement exact du trajet.

On voit cependant arriver de nouveaux moyens techniques de localisation permettant de tirer profit de ces informations détaillées (coordonnées exactes des points de départ et d'arrivée ; suivi détaillé du cheminement). Il sera en particulier intéressant de pouvoir tirer des enseignements de l'usage expérimental du GPS qui a été testé lors de la dernière enquête Transports (Insee-Inrets).

1.2. Les mesures basées sur des territoires

En pratique, on l'a dit, l'essentiel des statistiques de déplacements de personnes obtenues par des enquêtes sociodémographiques se réfèrent à des territoires. Plus exactement, les enquêtes - et les statistiques qui en découlent - ne se réfèrent pas explicitement à un objet géographique mais à un pouvoir administratif ou politique qui lui-même exerce sa compétence sur un territoire physiquement bien délimité. Pour certaines enquêtes, on ne se contente pas des limites administratives mais on construit des zonages d'étude qui peuvent être établis sur mesure (zones d'enquêtes des enquêtes déplacements) ou de façon générale (IRIS, zones d'emploi, aires urbaines, unités urbaines) en s'appuyant sur des données physiques ou socio démographiques. Ces zonages sont souvent bâtis -mais pas toujours- à partir de briques de base formées d'un territoire administratif. L'essentiel des informations issues de la statistique classique a donc comme support des territoires ce qui, on le verra, demande quelques précautions s'agissant de la mesure des flux et des calculs qui s'y rapportent.

Plusieurs enquêtes sur la population fournissent des informations du type « origine-destination ». La précision en est souvent de niveau communal : on connaît la commune d'origine et la commune de destination. C'est le cas en particulier des informations tirées du recensement de la population (par exemple les déplacements domicile-travail ou les migrations) ; c'est aussi le cas de la plupart des enquêtes portant sur les déplacements ainsi que celles sur les touristes. On a déjà évoqué³ comment cette situation est susceptible d'évoluer avec les changements profonds qui affectent les techniques de géolocalisation.

Ces enquêtes auprès de la population fournissent donc des résultats par zone. L'information la plus précise (coordonnées du point de départ et du point d'arrivée) est parfois recueillie par l'enquête : dans le cas, par exemple, d'une enquête au domicile, on connaît l'adresse de résidence, que l'on peut géolocaliser grâce à un GPS dont dispose l'enquêteur ou à partir d'un répertoire d'immeuble ou encore avec Google Earth ou tout autre outil de géolocalisation accessible gratuitement sur le net. Si on recueille également l'adresse du lieu de travail on peut donc disposer des coordonnées exactes des points de départ et d'arrivée. Ceci permet de calculer une distance mathématique (euclidienne, de Manhattan...) entre ces points. Cette distance calculée ne correspond en général pas exactement à la distance réellement parcourue par l'enquêté entre son domicile et son lieu de travail. Quelques rares

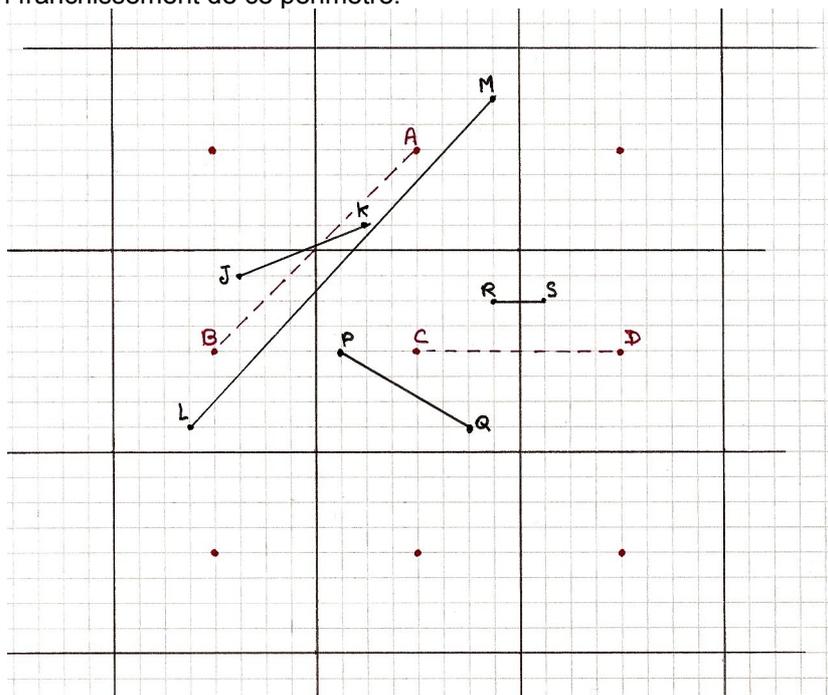
² Pour en savoir plus, voir en [2] le rapport de Marc Christine au Cnis ainsi que le dossier de l'enquête EVE (enquête sur les visiteurs de l'étranger) Ministère du Tourisme/Insee/Banque de France

³ voir au paragraphe 1.1.2. Les mesures basées sur des points

enquêtes permettent de recueillir cette information exacte, soit à l'aide d'un GPS confié à l'enquêté et qui enregistre le chemin parcouru⁴, soit à l'aide d'un SIG par le truchement duquel l'enquêté peut décrire exactement son parcours à l'enquêteur⁵. Dans le cas général on ne recueille – ou on ne retient – que l'information sur le territoire et non sur le point lui-même : on connaîtra par exemple en final la commune de résidence et la commune de lieu de travail. La distance retenue pour mesurer un déplacement domicile-travail sera donc la distance entre le chef-lieu de la commune de résidence et le chef-lieu de la commune de lieu de travail. Elle peut être calculée mathématiquement à partir des coordonnées des chefs-lieux ou tirée d'un référentiel kilométrique routier disponible dans le commerce.

1.2.1. Les calculs sur les territoires

Lorsque l'objet géographique de référence est un territoire, la base de travail est un plan maillé⁶ : chaque territoire forme une maille et l'ensemble des mailles forme une partition du plan. Il n'y a évidemment pas d'interstices entre les mailles et chaque point appartient à une maille et une seule. Le maillage utilisé peut être « naturel » (communes, départements...) et donc irrégulier ou bien formé de polygones réguliers (carrés, hexagones...). La distance retenue pour mesurer un déplacement est la distance entre le centre – au sens géométrique pour un polygone, au sens « chef-lieu » pour un territoire « naturel » - de la maille à laquelle appartient le point de départ et le centre de la maille à laquelle appartient le point d'arrivée. Par conséquent, tout mouvement effectué à l'intérieur du périmètre délimitant le territoire est considéré comme nul ; seuls sont pris en compte les mouvements occasionnant un franchissement de ce périmètre.



Commentaire sur le graphique : dans le maillage formé par les grands carrés, les déplacements entre les points J et K d'une part, L et M d'autre part seront tous deux assimilés à un déplacement entre A et B. Le déplacement entre les points P et Q est interne à la maille et donc non pris en compte. Le déplacement entre R et S sera assimilé à un déplacement entre C et D. On peut remarquer que les distances JK et LM, notablement différentes lorsqu'on les mesure de point à point, sont équivalentes dans ce référentiel maillé. La distance PQ est supérieure à RS lorsqu'on mesure de point à point et pourtant, en termes de distances maillées, la distance PQ est considéré comme nulle tandis que la distance RS vaut CD.

⁴Ce procédé a été expérimenté dans l'Enquête Transports réalisée par l'Insee en 2007-2008

⁵Procédé utilisé pour des enquêtes par téléphone réalisées par Affimétrie et destinées à mesurer la fréquence de passage devant des panneaux publicitaires

⁶On consultera avec profit le site pédagogique de Claude Grasland [6] et en particulier les chapitres « les mesures d'équité territoriale » et « les mesures de concentration territoriale »

Le mouvement peut être « entrant » lorsqu'un individu pénètre à l'intérieur du territoire ou « sortant » lorsqu'un individu quitte le territoire. Les flux « traversant » le territoire ne sont en général pas pris en compte sauf si cette traversée est marquée par un arrêt d'une certaine durée, par exemple un séjour d'au moins une nuitée dans le cas du tourisme. Les enquêtes renseignent en général sur le territoire de départ et le territoire d'arrivée mais pas sur l'itinéraire et donc pas sur les territoires traversés et encore moins sur les voies empruntées.

Cette distinction entre flux internes (« stables ») non pris en compte et flux externes (« entrées » ou « sorties ») seuls pris en compte est le facteur déterminant la mesure des flux entre les territoires.

Un biais peut venir de ce que les déplacements de courte distance sont pris en compte s'ils traversent la frontière mais ne le sont pas pour une distance équivalente parcourue à l'intérieur du territoire. Dans un précédent article déjà cité j'avais écrit : *« Plus la maille territoriale est fine, plus ce risque de biais est faible. L'incidence n'est négligeable que si l'essentiel des mouvements mesurés a une ampleur supérieure au diamètre de la surface. La taille de la maille territoriale a donc une incidence directe sur le volume de flux mesurés. Pour un même volume de flux, on n'obtiendra pas les mêmes chiffres selon le maillage territorial utilisé. »*

C'est pour vérifier cette assertion que j'ai entrepris la simulation qui est décrite plus loin. On pourra constater que les résultats ne confirment pas réellement le texte cité.

Si, de plus, on opère sur des mailles de tailles notablement différentes, le biais ne sera pas homogène. Dans ces conditions, il ne semble pas légitime de comparer des flux mesurés sur des maillages différents ou hétérogènes. C'est pourtant ce qui est souvent fait dans les comparaisons internationales où seule l'entité nationale est prise en compte. Cela revient à comparer dans un même classement des entités géographiques aussi disproportionnées que les USA dans leur ensemble et chacun des pays d'Europe pris isolément : - un voyage New-York/Los Angeles reste intérieur au pays et donc non comptabilisé ; - un voyage Paris/Bruxelles ne dure qu'une heure vingt minutes en TGV mais il est international et donc pris en compte. Ce problème, assez général en analyse territoriale⁷ revêt une importance particulière pour l'analyse des flux touristiques internationaux, comme cela a été montré dans le papier déjà cité. La justification de cette façon de faire est en réalité d'ordre financier. A partir du moment où un pays dispose d'un budget propre, la visite de ressortissants d'un autre pays -à condition qu'elle donne lieu à des dépenses sur le sol national du pays visité- se traduira par des flux financiers comptabilisés dans la Balance des paiements de chacun des deux pays.

1.2.2. Les changements d'échelle

Les principaux problèmes de calcul sur les flux mesurés entre des territoires se posent lorsque l'on veut changer d'échelle pour transformer une information obtenue sur un maillage territorial en une information sur un autre maillage. Cette opération est possible, moyennant certaines conditions, dans le sens ascendant (regroupement de petits territoires en un territoire plus vaste) mais elle n'est pas directement possible dans le sens descendant.

1.2.2.1. Passer d'une petite maille à une plus grosse maille

Le problème se pose, par exemple, quand on dispose de la mesure des flux entre communes et qu'on veut calculer des flux entre départements ou encore si on dispose de flux entre pays et qu'on veut calculer des flux entre continents.

Pour effectuer un tel regroupement, le calcul consistera à éliminer les flux internes aux nouvelles mailles et à recalculer par addition les flux entrants et sortants. Pour cela il faut considérer comme internes, et donc éliminer, les flux entre chacune des mailles initiales qui ont été regroupées pour former le nouveau territoire. Dans l'exemple du passage du niveau communal au niveau départemental, tous les flux mesurés entre les communes d'un même département deviennent internes au nouveau territoire qu'est le département et ne doivent donc plus être comptabilisés.

Puis on additionnera les flux sortants de chacune des mailles initiales vers une destination extérieure à ce nouveau territoire. Enfin on additionnera les flux entrants dans chacune des mailles initiales en provenance de l'extérieur du nouveau territoire.

Pour effectuer cette opération, il est donc nécessaire de disposer de l'information détaillée sur les relations entre chacune des mailles initiales. Au minimum il faut pouvoir distinguer parmi les flux initialement mesurés ceux qui deviennent internes aux nouvelles mailles et ceux qui restent externes. Si on ne dispose, pour chaque maille territoriale, que de la somme des flux entrants et de la somme

⁷Voir en [4] ...map generalisation of discrete social phenomena : Statistical problems... [Grasland *et alii*, 2000]

des flux sortants, il ne sera pas possible d'en déduire une information sur un regroupement territorial de niveau supérieur.

Le fait de regrouper des mailles pour former un territoire plus vaste a pour conséquence générale une diminution des flux mesurés puisqu'une partie de ces flux deviennent internes : la somme des flux entre départements sera toujours plus faible que la somme des flux entre communes ; la somme des flux entre continents sera toujours plus faible que la somme des flux entre les pays.

1.2.2.2. Passer d'une maille plus grande à une maille plus petite

Cette opération n'est pas directement possible par le calcul. Si on ne dispose pas de l'information détaillée au niveau de la maille la plus fine, on peut éventuellement chercher à modéliser ces flux en faisant des hypothèses, à partir d'éléments d'informations dont on pourrait disposer par ailleurs, sur la distance moyenne parcourue et la localisation privilégiée des émetteurs (population des lieux de résidence par exemple) ou des récepteurs (capacité d'hébergement des lieux touristiques par exemple). La fiabilité du résultat obtenu dépend alors de la qualité des informations complémentaires et des hypothèses retenues pour construire le modèle.

Ces difficultés liées à la taille des mailles territoriales et aux règles de calcul à appliquer pour des changements de maille sont source de nombreuses incertitudes -voire même d'erreurs manifestes- dans les évaluations existantes du nombre de touristes étrangers dans un pays donné et dans les chiffres publiés sur les flux internationaux de touristes dans le monde.

2. Étude de l'influence du maillage sur la mesure des flux

2.1. Le but de l'étude

On s'intéresse ici plus précisément à l'impact du découpage territorial sur les résultats obtenus dans la mesure des flux lorsque les objets géographiques de référence sont des territoires comme c'est le cas dans la plupart de nos enquêtes.

Nous avons précédemment supposé que le découpage de l'espace en territoires de différentes tailles avait une influence sur les résultats de la mesure. En particulier, puisque les déplacements ne sont pas pris en compte lorsqu'ils sont internes à un territoire, on pourrait logiquement s'attendre à ce que, plus grande est la taille de la maille, plus nombreux sont les déplacements ignorés car internes, ce qui devrait affaiblir d'autant la somme des distances parcourues mesurées par l'enquête.

Pour vérifier ce présupposé et analyser l'influence du maillage sur la mesure des flux, on a réalisé une simulation dont les grandes lignes sont les suivantes : après avoir choisi au hasard un point de départ A et un point d'arrivée B on mesure la distance, d'une part de point à point, d'autre part de maille à maille en utilisant différents maillages comme référentiels. Après avoir répété l'opération pour un grand nombre de couples de points, on compare les valeurs obtenues.

2.2. Les principaux résultats

On détaillera plus loin le détail des simulations réalisées, les paramètres testés ainsi que des pistes de formulation.

Les principaux résultats sont que :

Lorsque le référentiel est formé par des mailles carrées régulières et alignées et que les points de départ et d'arrivée sont équitablement répartis dans le plan, la somme des distances de Manhattan mesurées de maille à maille est en moyenne égale à la somme des distances de Manhattan mesurées de point à point, quelle que soit la taille de la maille.

Dans les mêmes conditions, le rapport entre la somme des distances euclidiennes mesurées de maille à maille et la somme des distances euclidiennes mesurées de point à point varie de 1 pour des petites mailles à 1,27 (soit $4/\pi$) pour des grandes mailles.

Lorsque le référentiel est formé par des mailles hexagonales, les résultats sont peu différents selon la métrique de distance utilisée et sont assimilables à ceux obtenus dans le cas d'un maillage de carrés alignés avec la distance euclidienne.

Une répartition non uniforme des points peut modifier fortement ces résultats : en particulier, lorsque les points sont situés au centre de la maille, plus la taille de la maille est grande et plus les petits déplacements sont ignorés dans les mesures de maille à maille.

3. La simulation

3.1. Le principe

Un point de départ est tiré au hasard dans le plan par génération aléatoire de ses coordonnées (x, y). Puis on choisit la direction du déplacement par tirage aléatoire d'un angle entre 0 et 360°. On choisit ensuite une distance par tirage aléatoire dans une table de distances équiprobables⁸. Tout ceci permet de déterminer les coordonnées d'un point d'arrivée à partir de celles du point de départ.

On dispose donc d'un couple de points (A, B) dont on connaît les coordonnées (xA, yA) et (xB, yB). Pour simuler un déplacement entre ces deux points il faut définir un chemin. On utilisera deux variantes, définies chacune par une métrique de distance : le chemin à vol d'oiseau (distance euclidienne), et le chemin en zigzag (distance de Manhattan⁹).

Pour un déplacement donné entre le point de départ A et le point d'arrivée B, la distance pourra être mesurée point à point ou selon un référentiel territorial. Différents référentiels sont utilisés dans cette simulation : ils sont constitués par des maillages réguliers pour lesquels on fait varier la taille de la maille de base. Plusieurs formes de mailles ont été testées : carrées ou hexagonales. Pour les mailles carrées on a également utilisé des agencements différents : carrés alignés ou carrés décalés (mur de briques).

On rappelle la règle de mesure des déplacements lorsqu'on utilise un référentiel territorial : si le point de départ et le point d'arrivée se trouvent dans la même maille, le déplacement n'est pas comptabilisé ; si le point de départ et le point d'arrivée sont dans des mailles différentes, on prend en compte pour le déplacement la distance séparant les centres des deux mailles. Toutes ces mesures sont effectuées selon les deux métriques : euclidienne et Manhattan.

Toutes les mesures, selon chacun des référentiels et selon chacune des métriques, sont effectuées simultanément pour chaque couple de points départ-arrivée de façon à ce que les résultats puissent être comparés.

Après avoir tiré au hasard un grand nombre de points et générés les déplacements à partir de ces points, on compare les sommes des distances parcourues selon qu'on les a mesurées point à point ou maille à maille, selon les différents référentiels.

3.2. Les règles retenues pour la simulation

On travaille sur un plan auquel on applique différents maillages de forme et de taille différentes.

Pour la forme on utilise des carrés alignés ou des hexagones.

Variante : carrés décalés, rectangles (alignés ou décalés), triangles

Lorsque les carrés ou les rectangles sont décalés, les côtés supérieurs des colonnes paires sont alignés sur le milieu des côtés latéraux des colonnes impaires (modèle du mur de briques).

Les hexagones sont par nature toujours décalés.

⁸ Chaque valeur de distance présente dans cette table a une probabilité égale d'être choisie. L'introduction de cette table de distances équiprobables permet de réaliser des simulations avec des données issues d'enquêtes donnant la répartition des déplacements selon la distance parcourue.

⁹ Pour deux points P1 (x1, y1) et P2 (x2, y2) : distance euclidienne DE = racine carrée [(x1-x2)**2 + (y1-y2)**2] et distance de Manhattan DM = valeur absolue (x1-x2) + valeur absolue (y1-y2)

Pour un grand nombre de couples de points P : Somme (DEij) / Somme (DMij) = 4 / π = 1,27

Pour la taille, on choisit la longueur c du côté du carré dans une liste (liste paramétrable).
Pour tester en parallèle différentes formes de mailles on leur donne la même surface ;
par extension, on appellera maille de taille c les mailles de surface c^2 , quelle que soit leur forme.

Le centre de la maille centrale est fixé à l'origine ($x=0, y=0$)

On détermine les coordonnées x et y d'un point de départ par tirage aléatoire dans un champ délimité (limites paramétrables).

A partir de ce point on génère un certain nombre de déplacements (nombre paramétrable) :

Pour cela on choisit une distance par tirage aléatoire dans une table de distances équiprobables¹⁰ (table paramétrable) et on choisit un angle par tirage aléatoire entre 0 et 360°

Partant du point de départ, on détermine les coordonnées du point d'arrivée selon le déplacement défini par l'angle et la distance euclidienne.

Variante : les coordonnées du point d'arrivée sont obtenues par un tirage aléatoire dans un champ délimité comme pour le point de départ. Les limites du champ sont paramétrables indépendamment des limites retenues pour le point de départ. Cette variante est supposée réduire les effets de bord (déplacements à partir du champ émetteur et dirigés vers l'extérieur du champ) mais, utilisée telle quelle, elle interdit de choisir les distances utilisées.

3.3. L'exercice de simulation

La simulation consiste à générer un grand nombre (paramétrable) de points de départ et à générer un certain nombre (paramétrable) de voyages à partir de chacun de ces points.

Pour cette simulation on se contente de voyages ne comprenant qu'un seul déplacement et on ne prend en compte que l'aller et non le retour. Donc ici les termes de « voyage » et de « déplacement » sont équivalents et s'appliquent au trajet simple effectué d'un point de départ à un point d'arrivée.

Pour chaque couple de points de départ et d'arrivée on détermine la distance point à point et la distance entre les centres des mailles auxquelles appartient chacun de ces deux points. Cette distance entre mailles est simultanément déterminée selon plusieurs tailles et plusieurs formes de mailles.

Pour le calcul des distances on utilise simultanément deux métriques différentes : la distance euclidienne (à vol d'oiseau) et la distance de Manhattan (en zigzag).

En fin de processus on dispose des sommes des distances calculées point à point d'une part et d'autre part entre mailles selon les différentes formes et tailles retenues pour ces mailles.

Pour chacune des combinaisons forme/taille, on dispose également du nombre de déplacements pris en compte, c'est à dire dont les points de départ et d'arrivée appartiennent à des mailles différentes. On dénombre également les déplacements considérés comme internes (points de départ et d'arrivée dans la même maille) et la somme des distances point à point correspondant à ces déplacements internes.

On calcule, pour chaque combinaison « forme de maille / métrique de distance » le rapport entre les distances entres points et les distances entre mailles et sa variation selon la longueur c du côté de la maille carrée.

Le programme¹¹ de simulation est écrit en langage Python (logiciel libre).

¹⁰ Chaque valeur de distance présente dans cette table a une probabilité égale d'être choisie. L'introduction de cette table de distances équiprobables permet de réaliser des simulations avec des données issues d'enquêtes donnant la répartition des déplacements selon la distance parcourue.

¹¹ Le code du programme est donné en annexe. Il est auto-documenté et décrit en particulier l'algorithme utilisé pour traiter le maillage en hexagones.

3.4. Les résultats de la simulation

3.4.1. Les résultats suivent une loi

Pour un protocole d'expérimentation donné (jeu de paramètres inchangé) les résultats obtenus par des simulations successives sont identiques dès lors que le nombre de voyages simulé est assez grand¹². Ceci montre que le résultat n'est pas aléatoire mais suit une loi. Dès lors on pourrait se demander s'il est bien légitime de chercher à montrer par des simulations ce qui pourrait l'être plus formellement par des calculs. En pratique, ces lois sont, dans le cas général, très complexes, ce qui justifie a posteriori une approche par simulation.

Quelques approches de la détermination de ces lois se trouvent dans la littérature. On pourra par exemple consulter la rubrique de Wikipedia consacrée au problème de l'aiguille de Buffon.

On donnera en annexe une formulation proposée par Jacques Choné (extension de Laplace du problème de l'aiguille de Buffon).

On se contentera ici d'une approche plus pragmatique : après avoir décrit les résultats obtenus par les simulations, on fera quelques développements sur leur justification.

3.4.2. Une relative invariance des distances selon le référentiel

La première conclusion des simulations est que, sous réserve de bonnes conditions, la somme des distances parcourues est à peu près égale selon qu'on la mesure entre mailles ou entre points.

Plus précisément, pour des mailles carrées alignées et en utilisant la distance de Manhattan, le rapport entre la somme des distances maille à maille et la somme des distances point à point oscille autour de 1 lorsqu'on fait varier la taille de la maille. L'ampleur de l'oscillation est très faible pour les plus petites mailles et s'accroît avec la taille des mailles.

Toujours pour des mailles carrées alignées, on obtient un résultat différent en utilisant la distance euclidienne. Le rapport entre la somme des distances maille à maille et la somme des distances point à point est toujours égal à 1 pour des petites mailles mais il augmente progressivement pour atteindre une valeur oscillant autour de 1,27 pour les très grandes mailles. Ce rapport connaît des oscillations du même ordre que pour la distance de Manhattan.

Pour des mailles hexagonales, les résultats diffèrent peu selon la métrique de distance utilisée. Le rapport est toujours de 1 pour les petites mailles et s'accroît progressivement avec la taille des mailles, comme les distances euclidiennes avec des mailles carrées alignées. Les oscillations sont de plus faible ampleur.

Pour que ces résultats soient vérifiés, deux conditions doivent être respectées :

- les points de départ et d'arrivée doivent être bien dispersés dans le plan
- les distances ne doivent pas être majoritairement petites par rapport à la maille

La deuxième condition semble la plus évidente : lorsque les distances sont petites par rapport à la maille, la probabilité est plus forte que les points de départ et d'arrivée soient situés dans la même maille et donc que les déplacements soient considérés comme internes.

En réalité la première condition est encore plus nécessaire. Le maximum d'effet est obtenu par une combinaison du non-respect des deux conditions.

Par exemple, lorsque tous les points de départ sont concentrés autour de l'origine, on est sûr que tous les déplacements inférieurs à la moitié du côté de la maille carrée seront internes à la maille centrée sur l'origine.

Par contre lorsque les points sont dispersés, la probabilité existe que des déplacements, même petits, puissent franchir au moins une limite de maille.

Une concentration proche d'une frontière de maille aura l'effet inverse d'une concentration au centre puisqu'elle augmentera la probabilité de franchissement des limites de maille et pourra donc augmenter de façon notable la somme des distances mesurées entre mailles lorsque les mailles sont grandes.

¹² Les tests courants ont été effectués avec 1 million de déplacements. Pour certains on a poussé jusqu'à 10 millions (mais le temps d'exécution devient important)

3.4.3.Ce qui ne reste pas invariant

Au fur et à mesure que l'on accroît la taille de la maille, le nombre de déplacements pris en compte diminue dans la mesure où de plus en plus de déplacements deviennent internes.

En conséquence, si la distance totale reste invariante la distance moyenne parcourue lors d'un déplacement augmentera puisqu'elle est calculée en rapportant la distance totale parcourue (qui est constante) au nombre de déplacements effectués (qui diminue).

En termes de résultats d'enquêtes, un maillage inadapté peut donc aboutir à des conclusions erronées : des individus pourront être considérés comme peu mobiles alors qu'ils se déplacent beaucoup mais sur de courtes distances. On pourra en déduire que certaines voies sont peu fréquentées alors qu'elles sont très fréquentées par beaucoup de gens qui parcourent de courtes distances.

3.4.4.Des variations selon la métrique de distance et la forme de la maille

Pour étudier les résultats des simulations, on analyse les variations, selon la taille de la maille, de l'indice Ind-D donnant le rapport entre la somme des distances D_m mesurées entre les mailles et la somme des distances D_p mesurées entre les points.

D_m = somme des distances mesurées en utilisant un référentiel défini par un maillage m ;

un maillage m est définie par la forme (carré, hexagone...) et la surface de ses mailles (égale à c^2 , ce qui correspond à un carré de côté c)

Conventions utilisées :

DV = distance euclidienne (Vol d'oiseau)

DZ = distance de Manhattan (Zigzag)

DV4 et DZ4 pour une maille carrée (4a pour des carrés alignés, 4d pour des carrés décalés)

DV6 et DZ6 pour une maille hexagonale

Pour des déplacements dont les distances sont uniformément réparties entre 0 et 1000 avec un champ émetteur dont les points de départ ont des abscisses et des ordonnées comprises entre -5000 et +5000, simulation réalisée avec 10 millions de voyages, sur des maillages carrés et hexagonaux avec c variant de 10 en 10 jusqu'à 5000 :

Les 4 indices Ind-DV4, Ind-DZ4, Ind-DV6 et Ind-DZ6 restent très près de 1 pour les valeurs de c de 1 à 100

Toutes les courbes laissent apparaître des tendances ondulatoires, l'amplitude des ondulations augmentant avec c .

Ind-Dv4 et Ind-DZ4 divergent légèrement : Ind-DZ4 oscillant autour de 1 tandis que Ind-DV4 monte au-dessus de 1

Ind-DV6 et Ind-DZ6 ont des variations très proches ; leur ondulation est moins prononcée que celle des Ind-D4 ; leur tendance est à la hausse légère, assez proche de celle de Ind-DV4

3.4.5.Quelques courbes obtenues par la simulation

Les deux graphiques présentés résument les différents types résultats obtenus.

3.4.5.1.Distances bien réparties

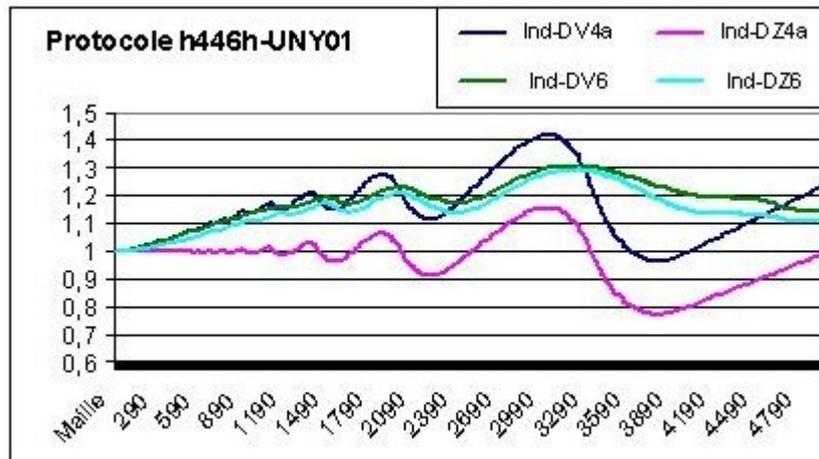
Dans le premier graphique, les distances des déplacements se répartissent sur une plage d'étendue suffisamment grande et les points de départ sont dispersés de façon aléatoire.

On y constate d'abord que, pour des petites mailles, la somme des distances mesurées de maille à maille est égale à la somme des distances mesurées de point à point, quelle que soit la forme de la maille et quelle que soit la métrique de distance utilisée. Ceci est évident, une mesure effectuée avec un maillage très fin étant sensiblement équivalent à une mesure de point à point.

On y constate ensuite que le rapport entre la somme des distances mesurées de maille à maille et la somme des distances mesurées de point à point oscille autour de 1 lorsqu'on utilise la distance de Manhattan avec un maillage à carrés alignés. L'ampleur des oscillations augmente avec la taille des mailles.

Avec la distance euclidienne et pour le même maillage carré, ce rapport part de la même valeur 1 mais s'en écarte pour osciller autour d'une valeur que l'on estime à 1.27. Les oscillations sont très semblables à celles constatées pour la distance de Manhattan.

L'utilisation d'un maillage hexagonal atténue les oscillations ; les résultats obtenus pour le rapport des distances selon qu'elles sont mesurées entre mailles ou entre points diffèrent peu d'une métrique à l'autre et semblent proches, aux oscillations près, de ceux obtenus avec la distance euclidienne dans un maillage à carrés alignés.

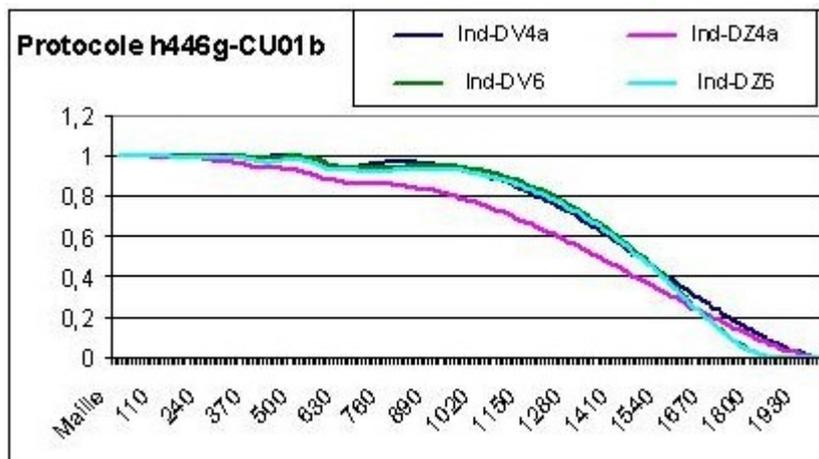


3.4.5.2. Distances faibles par rapport à la maille

Le second graphique montre une simulation où les distances ont été choisies dans une plage limitée tandis que tous les points de départ des déplacements sont proches de l'origine représentée par le point (0,0) qui a été fixé comme centre pour tous les maillages. Au delà d'une certaine taille de maille, le nombre de déplacements éliminés dans la mesure des distances de maille à maille car internes à une maille devient non négligeable, puis majoritaire.

Dans ce cas les distances mesurées de maille à maille tendent vers zéro et donc le rapport entre la somme des distances mesurées de maille à maille et la somme des distances mesurées de point à point tend également vers zéro.

Ce résultat est à peu près équivalent quelle que soit la forme des mailles et quelle que soit la métrique de distance utilisée.



4. Éléments de formalisation

L'intuition suggère que, plus la taille des mailles est importante et plus la somme totale des distances mesurées de maille à maille est diminuée par l'élimination des déplacements internes à la maille. Le résultat des simulations montre qu'il n'en est rien : la somme des distances mesurées de maille à maille reste à peu près constante quelle que soit la taille de la maille. Elle demeure égale - ou même supérieure dans certains cas - à la somme des distances mesurées de point à point.

On peut en faire une démonstration assez simple pour les déplacements effectués le long d'une droite. On montrera ensuite que dans le plan les choses se compliquent. On examinera en premier lieu le cas le plus simple des mailles carrées combinées à la distance de Manhattan et on extrapolera aux mailles carrées avec la distance euclidienne.

On trouvera par ailleurs en annexe une formulation plus élaborée proposée par Jacques Choné.

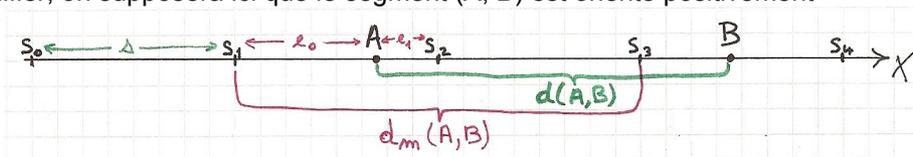
4.1. Déplacements le long d'une droite segmentée

Soit une droite D divisée en segments égaux de longueur s

On nomme $S_0, S_1, S_2, \dots, S_n$ les origines des segments de rang 0, 1, 2...n

Soit sur cette droite, deux points quelconques A et B

pour simplifier, on supposera ici que le segment (A, B) est orienté positivement



La distance, mesurée de point à point entre A et B est $d(A,B) = x_B - x_A$

$d(A,B)$ peut se décomposer en un nombre entier n de segments de longueur s , et un reste r :

$$d(A,B) = n * s + r$$

remarque : si $[d(A,B) < s]$ alors $n = 0$

On appellera distance entre mailles $dm(A,B)$ la distance entre les centres des segments auxquels appartiennent respectivement le point A et le point B (remarque : la distance entre les centres des segments est équivalente à la distance entre les origines)

le point A est positionné sur le segment de rang a délimité par les points S_a et S_{a+1}

le point B est positionné sur un segment de rang b délimité par les points S_b et S_{b+1}

La distance entre mailles est déterminée par le nombre de segments qui séparent S_a de S_b

$$dm(A,B) = (b - a) * s$$

il faut prendre en compte la position du point A sur son segment :

on appelle e_0 la distance(S_a, A) et e_1 la distance(A, S_{a+1})

si $r < e_1$ alors $b = a + n \quad \rightarrow b - a = n \quad \rightarrow dm(A,B) = n * s \quad \rightarrow dm(A,B) = d(A,B) - r$

si $r \geq e_1$ alors $b = a + n + 1 \quad \rightarrow b - a = n + 1 \quad \rightarrow dm(A,B) = (n + 1) * s \quad \rightarrow dm(A,B) = d(A,B) - r + s$

cas particulier : si $[d(A,B) < s]$ et si $[d(A,B) < e_1]$ alors les points A et B sont situés sur le même segment et la distance entre mailles est nulle

On voit que la distance entre mailles est égale à la distance entre points, systématiquement amputée de r et augmentée de s dans le cas où $r \geq e_1$

Si l'on ne fixe aucune condition sur la position de A alors les valeurs de e_1 se répartissent de façon équiprobable dans l'intervalle $(0, s)$

Si l'on ne fixe pas de condition sur la longueur du segment (A, B) alors les valeurs de r se répartissent de façon équiprobable dans l'intervalle $(0, s)$

On en déduit :

La valeur moyenne de r est $s/2$

La probabilité que r soit $\geq e_1$ est $1/2$

Par rapport à la distance point à point, la distance entre-mailles est donc systématiquement amputée de r qui vaut en moyenne $s/2$ et elle est en moyenne une fois sur 2 augmentée de s

Donc en moyenne, pour un nombre important de couples de points quelconques, la somme des distances mesurées de point à point est égale à la somme des distances mesurées de maille à maille.

4.2. Déplacements dans un plan maillé

Sur le principe général, on pourrait penser qu'il n'y a pas de raison fondamentale pour que l'égalité des distances entre points et entre mailles, établie sur une droite segmentée, ne puisse être généralisée à un plan. Mais pour être vérifiée, cette égalité nécessite une équiprobabilité de la répartition des points sur un référentiel segmenté de façon uniforme. En pratique, aucune forme de polygone régulier utilisée pour mailler un plan ne permet d'obtenir une segmentation uniforme selon tous les angles des droites qui coupent ce maillage.

4.2.1. Maillage réalisé avec des carrés

Le cas de la maille carrée est particulièrement intéressant car c'est le plus facilement utilisable pour les calculs.

Soit un maillage formé de carrés de côté c superposés et alignés, délimités par des droites parallèles aux axes X et Y et dont le centre de la maille origine est $(0, 0)$.

4.2.1.1. Distances entre mailles selon la métrique

Dans la métrique de Manhattan, les distances entre les centres de mailles varient par paliers réguliers de pas c . Cette distance vaut c pour le voisinage de rang 1 formé des carrés dont les côtés sont adjacents à la maille carrée contenant le point de départ. La distance est de $2c$ pour le voisinage de rang 2 formé des carrés dont les côtés sont adjacents aux mailles du voisinage de rang 1. Elle est de $n \cdot c$ pour les carrés de voisinage n .

Dans la métrique euclidienne, on ne retrouve pas cette répartition régulière. La variation est toujours régulière et de pas c pour les carrés alignés horizontalement ou verticalement. Pour les autres mailles, la distance est fonction de l'angle formé par la droite qui relie le centre de la maille d'arrivée au centre de la maille de départ.

Prenons comme origine la maille de coordonnées $(0, 0)$

Les carrés adjacents horizontalement, d'abscisse 0, formant la couche de rang 0, sont séparés de la maille origine par des distances $c, 2c, 3c$, etc.

Considérons maintenant la couche horizontale de rang 1, formée des carrés d'abscisse c :

- pour le carré adjacent verticalement à la maille d'origine, de centre $(0, c)$ la distance est c
- pour le premier vers la droite, de centre (c, c) , la distance est égale à $c \cdot \sqrt{2}$
- pour le deuxième, de centre $(2c, c)$, la distance est égale à $c \cdot \sqrt{5}$
- pour le troisième, de centre $(3c, c)$, la distance est égale à $c \cdot \sqrt{10}$
- pour le n ème, de centre (nc, c) , la distance d est égale à $c \cdot \sqrt{1+n^2}$

Considérons la couche horizontale de rang 2, formée des carrés d'abscisse $2c$:

- pour le carré situé à la verticale de la maille d'origine, de centre $(0, 2c)$ la distance est $2c$
- pour le premier vers la droite, de centre $(c, 2c)$, la distance est égale à $c \cdot \sqrt{5}$
- pour le deuxième, de centre $(2c, 2c)$, la distance est égale à $c \cdot 2\sqrt{2}$
- pour le troisième, de centre $(3c, 2c)$, la distance est égale à $c \cdot \sqrt{13}$
- pour le n ème, de centre $(nc, 2c)$, la distance d est égale à $c \cdot \sqrt{4+n^2}$

Pour la couche de rang r , les distances seront :

- pour le carré situé à la verticale de la maille d'origine, de centre $(0, rc)$ la distance est rc
- pour le n ème, de centre (nc, rc) , la distance d est égale à $c \cdot \sqrt{r^2+n^2}$

On peut donc constater que l'exercice de calcul des distances entre mailles carrées n'a pas la même complexité selon que l'on utilise la distance de Manhattan ou la distance euclidienne.

4.2.1.2. Du point à la maille ; des complexités différentes selon la métrique

Cet écart de complexité selon la métrique choisie s'amplifie encore lorsqu'on entreprend de comparer les distances mesurées de point à point aux distances mesurées de maille à maille.

Dans la métrique euclidienne, les points séparés de l'origine par une distance d sont situés sur un cercle de rayon d ayant l'origine pour centre.

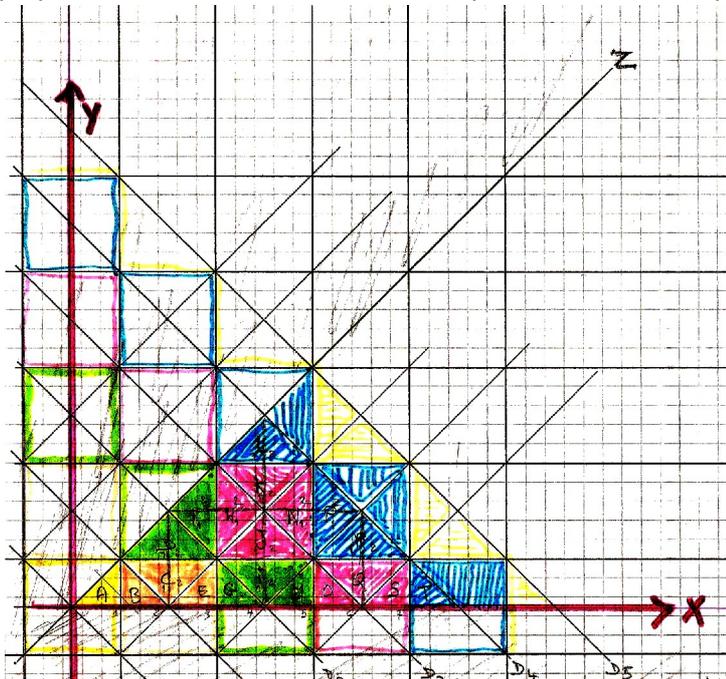
Dans la métrique de Manhattan, les points séparés de l'origine par une distance d sont situés sur un carré de côté $d/\sqrt{2}$ dont les axes x et y forment les diagonales ou, exprimé autrement, un carré de diagonale $2d$ dont les côtés sont inclinés de 45° par rapport à l'horizontale et à la verticale.

Choisissons l'origine (0, 0) comme point de départ. On cherche à comparer les distances D_p (mesurées directement entre les points) qui séparent ce point de départ des différents points d'arrivée dans le plan et les distances D_m mesurées entre les centres des mailles auxquelles appartiennent les points de départ et d'arrivée dans le maillage formé par des carrés de côté c . Puisque l'on a choisi le point origine comme point de départ, on peut se contenter, compte tenu des symétries, d'examiner le premier demi-quadrant, soit la portion de plan délimitée par l'axe des X et la droite Z inclinée à 45° par rapport à cet axe et bissectrice du premier quadrant. Tous les points d'arrivée étant possibles, on peut considérer que, dans ce demi-quadrant, la surface délimitée par une courbe d'équidistance d (cercle pour Euclide et carré pour Manhattan) représente la probabilité d'existence d'un déplacement de distance comprise entre 0 et d . On la note S_d . La surface de chacune des intersections de ce fuseau avec chacun des carrés du quadrillage donnera la probabilité que la distance de maille à maille prenne la valeur affectée à ce carré.

4.2.1.3. Du point à la maille avec la distance de Manhattan

Dans la métrique de Manhattan :

Dans le demi-quadrant considéré, les points équidistants du point de départ sont situés sur le segment de droite perpendiculaire à la bissectrice Z et coupant l'axe des X au point $x=d$.



Pour une valeur de D_p égale à c , la surface S_c est égale au quart de la surface d'un carré du maillage. Elle se partage à parts égales entre le carré origine (distance = 0) et le carré voisin (de distance = c). Ce qui signifie que, pour des distances entre points comprises entre 0 et c , il y aura équivalence en moyenne entre les distances mesurées de point à point et les distances mesurées de maille à maille. Si l'on examine plus finement ce qui se passe lorsque D_p varie de 0 à c : S_d est entièrement comprise dans le carré origine (distance = 0) tant que d est inférieur à $c/2$; S_d se partage ensuite entre le carré origine et le carré suivant (distance = c).

On peut noter que la variation n'est pas régulière : pour d variant de 0 à $c/2$, la surface de l'intersection IN_0 avec le carré origine croît régulièrement de 0 à $c/16$; pour d variant de $c/2$ à c la surface IN_0 croît de $c/16$ à $c/8$, le gain de surface étant maximum au début et minimum à la fin tandis que la surface IN_1 varie de 0 à $c/8$, soit un gain en surface double de celui de IN_0 , ce gain étant minimal au début et maximal en fin.

On peut renouveler l'exercice en augmentant d : on constate que l'on atteint toujours un équilibre pour les valeurs de d multiples de c mais que cet équilibre est une combinaison de déséquilibres pour les valeurs intermédiaires.

Cet exercice a été réalisé en prenant le centre du carré origine comme point de départ. On extrapolera en considérant que, en moyenne, on obtient un résultat semblable lorsqu'on déplace ce point de départ de façon uniforme dans l'ensemble du carré.

On a donc, dans le cas de la distance de Manhattan appliquée à un maillage carré, l'explication du résultat obtenu par la simulation : pour une distribution uniforme des points, le rapport entre les distances mesurées de point à point et les distances mesurées de maille à maille est en moyenne égal à 1 avec des valeurs oscillant autour de 1 selon des variations dissymétriques.

4.2.1.4. Du point à la maille avec la distance euclidienne

Dans la métrique euclidienne :

Dans le demi-quadrant considéré, les points équidistants du point de départ sont situés sur l'arc de cercle de rayon d .

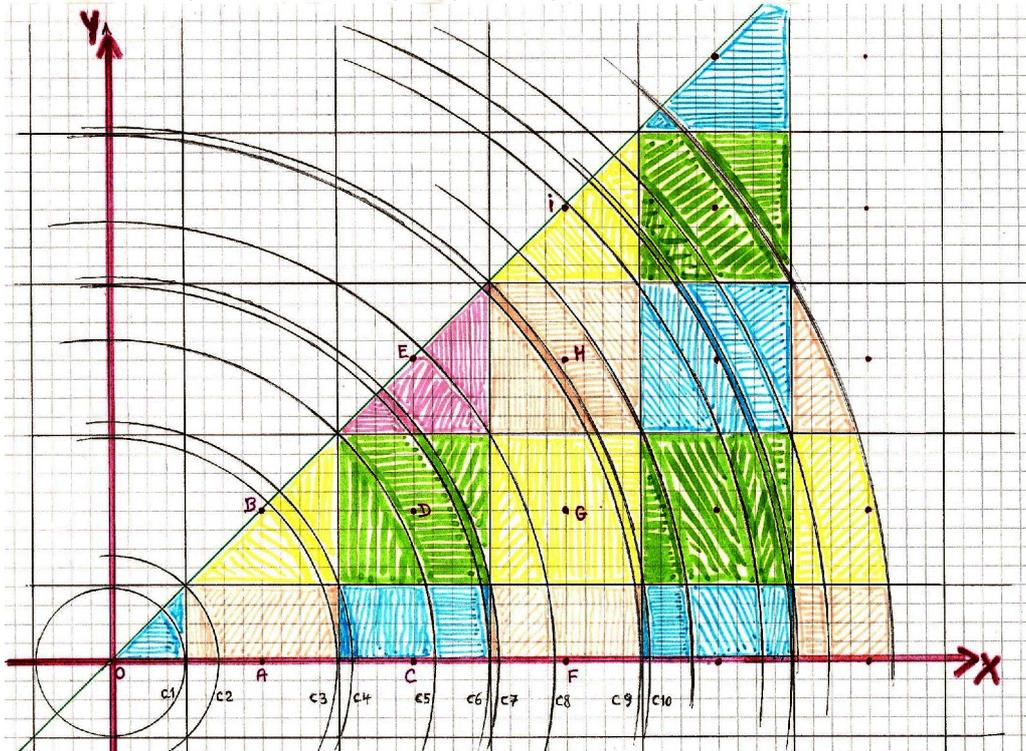
Pour D_p compris entre 0 et $c/2$ tous les points sont dans la même maille et $D_m = 0$

Pour D_p entre $c/2$ et $c\sqrt{2}$ alors $D_m = 0$ ou c selon l'angle

Pour D_p entre $c\sqrt{2}$ et $c\sqrt{3}/2$ alors $D_m = c$ ou $c\sqrt{2}$ selon l'angle

Pour D_p entre $c\sqrt{3}/2$ et $c\sqrt{3}/2$ alors $D_m = c\sqrt{2}$ ou c^2 ou $c\sqrt{5}$ selon l'angle

On peut continuer l'exercice, il se complexifie au fur et à mesure que l'on augmente les distances ; on en tirera la conclusion que le résultat ne peut pas s'exprimer de façon simple.



Les carrés du maillage étant délimités par des séries de parallèles aux axes X et Y, considérons une droite traversant ce maillage selon un angle A

Elle sera segmentée de façon régulière par les parallèles horizontales.

Elle sera segmentée de façon régulière par les parallèles verticales

Elle ne sera en général pas segmentée de façon régulière par les carrés formant le maillage

L'irrégularité de la segmentation est fonction de 2 facteurs : l'angle A que fait la droite avec l'axe des X et la position du point où elle coupe l'axe des X (que l'on appellera point origine)

Pour un angle de modulo 45° (0° , 45° , 90° , etc. ...) la segmentation sera régulière quel que soit le point origine.

Cette propriété sera vérifiée dans les cas où la segmentation n'est déterminée que par un seul des ensembles de parallèles (les horizontales ou les verticales), l'intersection avec leurs perpendiculaires se situant toujours à une intersection (c'est à dire à un coin d'un carré formant le maillage).

Cette condition est vérifiée, lorsque le point origine est situé sur une limite d'un carré formant le maillage, pour les angles A tels que tangente (A) (c'est à dire x/y) soit un rapport d'entiers.

exemples : $A = 26,6^\circ$ ($tg=1/2$) ; $A = 18,4^\circ$ ($tg=1/3$) ; $A = 14^\circ$ ($tg=1/4$) ; $A = 11,3^\circ$ ($tg=1/5$) et leurs complémentaires par rapport à 90° :

$A = 63,4^\circ$ ($tg= 2$) ; $A = 71,6^\circ$ ($tg= 3$) ; $A = 76^\circ$ ($tg= 4$) ; $A = 78,7^\circ$ ($tg= 5$)

Lorsque l'on fait varier l'angle et la position d'origine, on observe une irrégularité des segments ; cette irrégularité peut selon les cas, suivre une séquence plus ou moins complexe (exemple simple : un segment de longueur L1 suivi de 2 segments de longueurs L2 suivis d'un segment de longueur l1, etc.)

Pour un maillage du plan par des carrés de côté c il ne sera donc pas possible de disposer d'une segmentation régulière qui soit la même selon tous les angles. La question est de savoir si et comment toutes ces irrégularités finissent par se compenser. Le rapport entre Dm (somme des distances maille à maille) et Dp (somme des distances point à point) répondra donc à une loi complexe traduisant l'équilibre entre ces irrégularités. On ne tentera pas ici l'exercice de formulation de ces équilibres mais on donnera en annexe des éléments de pistes de tentatives de formulation. Nous nous contenterons ici des résultats de la simulation.

4.2.1.5. De la distance de Manhattan à la distance euclidienne

On a vu que les résultats obtenus en utilisant la distance de Manhattan étaient assez simplement explicables tandis que l'usage de la distance euclidienne rendait la formulation d'une complexité plus inaccessible. On peut cependant utiliser les résultats obtenus avec la distance de Manhattan pour aller plus loin dans les résultats obtenus avec la distance euclidienne.

Considérons comme acquis et suffisamment explicite le résultat obtenu en utilisant la distance de Manhattan avec un maillage formé de carrés réguliers et alignés : le rapport entre les distances mesurées de point à point et les distances mesurées de maille à maille est en moyenne égal à 1.

Une propriété connue est que, pour un grand nombre de couples de points, le rapport entre la somme des distances de Manhattan et la somme des distances euclidiennes est égal à $4/\pi$ soit 1,27

On considérera que cette propriété est vérifiée pour les distances mesurées de point et, par extension, pour les distances mesurées de maille à maille pour de petites mailles.

Lorsque la taille des mailles augmente jusqu'à devenir suffisamment grande pour que les distances entre les points puissent être considérées comme petites par rapport au côté de la maille. Alors les seuls déplacements qui franchiront une frontière de maille, et donc qui seront pris en compte dans les distances maille à maille, sont ceux qui émanent d'un point proche d'une frontière et rejoignent un autre point de l'autre côté de la frontière. Compte tenu des valeurs respectives de la longueur du côté de la maille et de la longueur du déplacement, l'essentiel des déplacements pris en compte relieront une maille à sa voisine immédiate. A part quelques « fuites » négligeables dans les coins du carré, il s'agira donc de liaisons entre carrés dont les côtés sont adjacents. Or, pour des carrés de côté adjacents, nous avons vu que la distance de maille à maille est la même dans les deux métriques de distances, euclidienne ou Manhattan, égale à la longueur du côté du carré.

On peut donc en déduire que, pour des petites mailles, la somme des distances de Manhattan est 1.27 fois supérieure à la somme des distances euclidiennes tandis que, pour des grandes mailles, les sommes des distances seront égales dans les deux métriques.

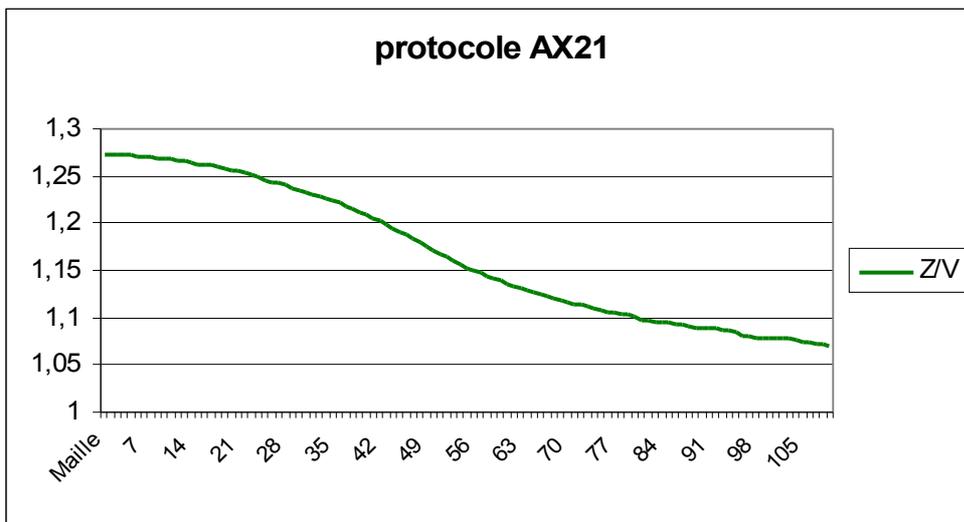
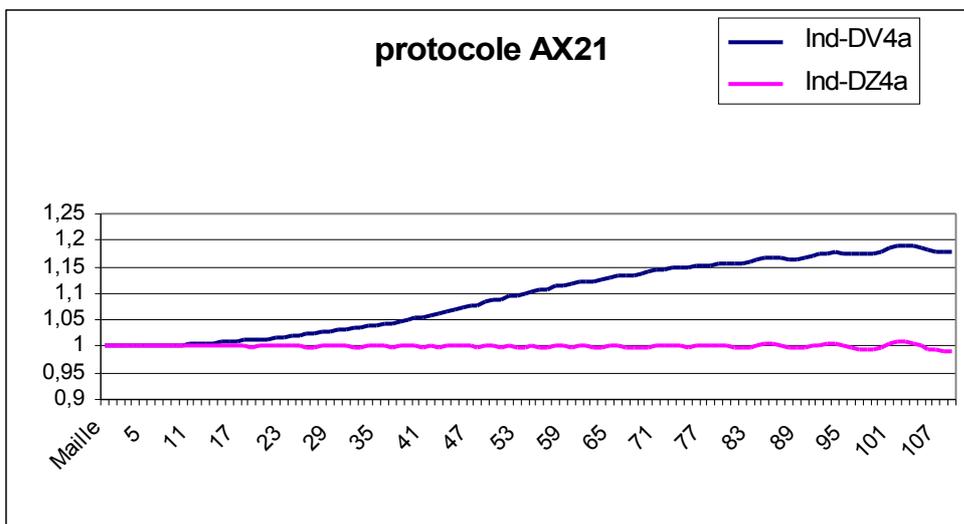
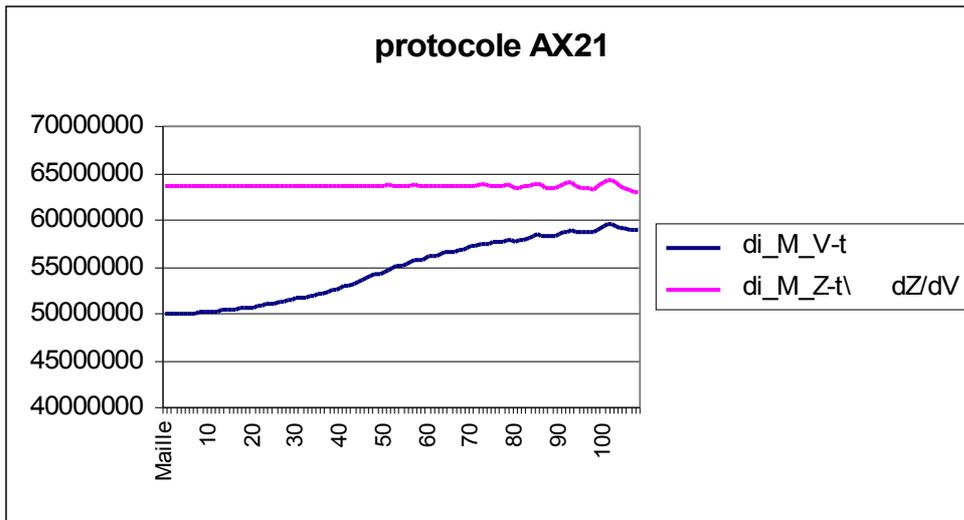
Ceci explique le résultat obtenu par simulation : pour des distances euclidiennes, le rapport entre la somme des distances mesurées de maille à maille et la somme des distances mesurées de point à point varie de 1 pour les petites mailles à 1.27 pour les grandes mailles.

Graphiques illustrant le rapport entre les distances mesurées maille à maille selon la métrique de Manhattan (Z comme zigzag) ou euclidienne (V pour vol d'oiseau) lorsque la taille de la maille varie.

Graphique-1 : somme des distances maille à maille

Graphique-2 : indice (somme des distances maille à maille)/(somme des distances point à point)

Graphique-3 : rapport entre distances maille à maille : Manhattan / euclidienne



4.2.2. Maillage réalisé avec des hexagones

Parmi les autres polygones réguliers permettant de réaliser un maillage du plan figure l'hexagone.

L'hexagone semble présenter une particularité intéressante pour un maillage : la distance entre le centre d'un hexagone et le centre de chacun des hexagones de son voisinage au premier degré est constante (égale à $c\sqrt{3}$ pour des hexagones de côté c). Mais cette régularité disparaît dès le deuxième degré de voisinage (2^{ème} couche) puisque les distances entre centres peuvent être soit $c\sqrt{3}$ soit $c\sqrt{2}$.

Pour un maillage hexagonal, pas plus que pour un maillage carré, on ne peut obtenir une segmentation qui soit régulière et indépendante de l'angle.

Le maillage hexagonal a néanmoins également été testé dans la simulation. Les résultats obtenus présentent moins d'irrégularités oscillatoires qu'avec un maillage carré. Les résultats diffèrent peu avec la métrique (Euclidienne ou Manhattan) et semblent plus proches de ceux obtenus pour un maillage carré avec la distance euclidienne.

Bibliographie

- [1] BERNADET M., 2003, *Les dispositifs statistiques d'observation de l'activité touristique au niveau national : présentation et évaluation*, Edition Observatoire National du Tourisme
- [2] CHRISTINE M., VASSILLE L., 2004, « Les enquêtes au bord des routes dans les domaines des transports et du tourisme : quelles alternatives ? » *Rapport CNIS* (Conseil National de l'Information Statistique) n° 82, février 2004, 64 pages
- [3] DEVILLE JC., MAUMY M., 2005 « Extension de la méthode d'échantillonnage indirect et son application aux enquêtes dans le tourisme », *Journées de Méthodologie Statistique de l'INSEE*, mars 2005
- [4] GRASLAND C., MATHIAN H., VINCENT J.M., 2000, « Multiscalar Analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences », *Statistical Journal of the United Nations ECE*, 17, IOS Press, 1-32
- [5] GRASLAND C., 2007, « Global city revisited by Tobler's first law of geography », *15th European Colloquium on Theoretical and Quantitative Geography 7-11, 2007*, Montreux, Switzerland
- [6] GRASLAND C., Site pédagogique <http://grasland.script.univ-paris-diderot.fr/> et en particulier les chapitres "Les mesures d'équirépartition" et "Les mesures de concentration spatiale" http://grasland.script.univ-paris-diderot.fr/anspa/concent/Chap1_1.htm
- [7] TERRIER C., SYLVANDER M., KHIATI A., MONCERE V., 2005, « Population présente : méthodes de détermination à partir des enquêtes sur les touristes. » Communication aux *Journées de Méthodologie Statistique de l'INSEE* – 15 mars 2005
- [8] TERRIER C. (dir.), 2006, « Mobilité touristique et population présente – Les bases de l'économie présentielle des départements », édition Direction du Tourisme, Version interactive accessible sur le site de Christophe Terrier <http://www.christophe-terrier.com> à la rubrique « Présence »
- [9] TERRIER C., 2006, « Flux et afflux de touristes : les instruments de mesure, la géomathématique des flux », in revue FLUX, Cahiers scientifiques internationaux Réseaux et Territoires, n° 65, juillet-septembre 2006

5. Annexes

5.1. Les algorithmes utilisés

Le programme Python utilisé pour ces simulations peut être fourni sur demande.

5.1.1. Détermination de l'appartenance d'un point à un hexagone du maillage

L'algorithme qui a été développé pour cette simulation pouvant être réutilisé pour d'autres travaux, il a paru utile d'en expliciter les grandes lignes. Le texte ci-dessous est un extrait de la documentation accompagnant le programme Python.

```
#---- Traitement du maillage hexagonal -----
#Le maillage est constitué d'hexagones de côté c
# rappel : pour l'hexagone, le rayon est égal au côté ; la hauteur (ou apothème)  $h=c*(\text{rac}(3)/2)$ 
#l'hexagone central a pour centre le point origine (0,0)
#les côtés supérieurs et inférieurs des hexagones sont horizontaux
#les autres côtés forment des angles de  $60^\circ$  ou de  $120^\circ$  avec l'horizontale
#Conventions :
#les droites D sont des parallèles faisant un angle de  $60^\circ$  avec l'axe des x (donc allant vers la droite pour y croissant)
#les droites G sont des parallèles faisant un angle de  $120^\circ$  avec l'axe des x (donc allant vers la gauche pour y croissant)
#les parallèles sont espacées de h (h= hauteur des triangles se joignant au centre de l'hexagone ; appelé apothème)
#les droites D et G coupent l'axe des y à des intervalles de 2h
#les droites D et G coupent l'axe des x à des intervalles de c
#on identifie les parallèles par leur rang :
#le rang 0 est affecté aux droites D et G qui passent au point d'origine (0,0)
#le rang 1 est affecté aux droites D et G qui coupent l'axe des y en (0,2h)
# la droite D1 coupe l'axe des x en (-c,0) ;
# la droite G1 coupe l'axe des x en (+c,0) ;
# D(rd) :  $y = \text{rac}(3)*x + rd*(2*h)$ 
# G(rg) :  $y = -\text{rac}(3)*x + rg*(2*h)$ 
#La bande de rang r est la partie du plan comprise entre la parallèle de rang r (inclue) et celle de rang r+1 (exclue)
#L'intersection d'une bande G et d'une bande D forme un losange de diagonale verticale 2h et de diagonale horizontale c
#Le sommet inférieur de ce losange est l'intersection des droites G(rg) et D(rd) :
# intersection [G(rg),D(rd)] --> I(xi,yi) avec  $xi=(c/2)*(rg-rd)$  et  $yi=h*(rg+rd)$ 
#
# Position d'un point P(xp,yp) dans un maillage hexagonal
# (recherche du centre de l'hexagone auquel appartient P)
#-déterminer à quelle bande D(rd) appartient P :
#  $rd = \text{math.floor}((y - \text{rac}(3)*x)/(2*h))$ 
#-déterminer à quelle bande G(rg) appartient P :
#  $rg = \text{math.floor}((y + \text{rac}(3)*x)/(2*h))$ 
#le losange déterminé par l'intersection des bandes Drd et Grg peut occuper 3 types de positions :
# 1) le losange est central : une de ses moitiés appartient à l'hexagone inférieur et l'autre à l'hexagone supérieur
# 2) le losange forme la partie droite d'un hexagone du maillage
# 3) le losange forme la partie gauche d'un hexagone du maillage
# Le choix entre l'une de ces 3 positions est déterminé par l'écart (modulo 3) entre les rangs rd et rg
#  $\text{ecart} = (rd - rg) \% 3 = 0 \text{ ou } 1 \text{ ou } 2$ 
# si = 0 : position centrale ;
```

```

# si abs(yp-yi)< h le point appartient à l'hexagone inférieur de centre=
intersection[G(rg),D(rd)]
# si abs(yp-yi)>= h le point appartient à l'hexagone supérieur de centre=
intersection[G(rg+1),D(rd+1)]
# si = 1 : position droite --> le centre de l'hexagone est l'intersection [G(rg),D(rd+1)]
# si = 2 : position gauche --> le centre de l'hexagone est l'intersection [G(rg+1),D(rd)]
#
# point P(xp,yp) ;
# intersection[G(rg),D(rd)] = I(xi,yi) = coordonnées du sommet inférieur du losange intersection des
bandes Grg et Drd
#

```

5.2. Les formulations mathématiques par Jacques Choné

5.2.1. Problème 95

Problème n°95 (proposé par Loïc Terrier)

On lance un spaghetti de longueur d sur un sol carrelé, les carreaux étant des carrés de côté unité. Quelle est la probabilité que le spaghetti soit à l'intérieur d'un des carreaux ?

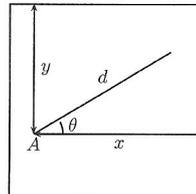
Problème n°95 (proposé par Loïc Terrier)

On lance un spaghetti de longueur d sur un sol carrelé, les carreaux étant des carrés de côté unité. Quelle est la probabilité que le spaghetti soit à l'intérieur d'un des carreaux ?

Solution par Jacques Choné (Clermont-Ferrand)

Il s'agit de l'extension de Laplace du problème de l'aiguille de Buffon. Voir les références [1] et [2].

Soit A l'extrémité gauche du spaghetti (on peut négliger le cas où il est vertical, cas de probabilité nulle) et C le carré contenant A (on peut négliger le cas où A est sur la frontière d'un carré, cas de probabilité nulle). Soit x (resp. y) la distance de A à la verticale de droite de C (resp. à l'horizontale du haut de C) et θ l'angle du spaghetti avec l'horizontale.



On suppose que les variables aléatoires x, y, θ sont indépendantes et que x et y suivent la loi uniforme sur $[0, 1]$ et θ la loi uniforme sur $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Soit E_h (resp. E_v) l'événement "le spaghetti rencontre une droite horizontale (resp. verticale)".

Cas où $d \leq 1$

On a :

$$\begin{aligned}
 P(E_h / \theta < 0) &= P(y < d \sin \theta / 0 < \theta) = \frac{1}{\frac{\pi}{2} \cdot 1 \cdot 1} \int_{x=0}^{x=1} \int_{\theta=0}^{\theta=\frac{\pi}{2}} \int_{y=0}^{y=d \sin \theta} dy \cdot d\theta \cdot dx \\
 &= \frac{1}{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} d \sin \theta \cdot d\theta = \frac{2d}{\pi}
 \end{aligned}$$

Pour des raisons de symétrie, on a également $P(E_h / \theta < 0) = \frac{2d}{\pi}$. (voir un calcul direct dans l'appendice a). On obtient alors :

$$P(E_h) = P(E_h / \theta < 0) \cdot P(\theta < 0) + P(E_h / \theta > 0) \cdot P(\theta > 0) = 2 \cdot \frac{2d}{\pi} \cdot \frac{1}{2} = \frac{2d}{\pi}.$$

Pour des raisons de symétrie, on a également $P(E_v) = \frac{2d}{\pi}$.

On a :

$$\begin{aligned}
 P(E_h \cap E_v / \theta < 0) &= P(y < d \sin \theta \text{ et } x < d \cos \theta / 0 < \theta) \\
 &= \frac{1}{\frac{\pi}{2} \cdot 1 \cdot 1} \int_{\theta=0}^{\theta=\frac{\pi}{2}} \int_{x=0}^{x=d \cos \theta} \int_{y=0}^{y=d \sin \theta} dy \cdot dx \cdot d\theta = \frac{1}{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} d^2 \sin \theta \cdot \cos \theta \cdot d\theta = \frac{d^2}{\pi}
 \end{aligned}$$

Comme ci-dessus, on a pour des raisons de symétrie $P(E_h \cap E_v / \theta > 0) = \frac{d^2}{\pi}$ et on en déduit

$$P(E_h \cap E_v) = \frac{d^2}{\pi}.$$

La probabilité demandée est donc :

$$1 - P(E_h \cup E_v) = 1 - P(E_h) - P(E_v) + P(E_h \cap E_v) = \boxed{1 - \frac{d(4-d)}{\pi}}.$$

5.2.2.Problème 95-bis

Problème n°95bis (proposé par Loïc Terrier)

On lance un spaghetti de longueur d sur un sol carrelé, les carreaux étant des carrés de côté unité. Quelle est l'espérance de la distance, X_d , entre les centres des deux carreaux où se trouvent les extrémités du spaghetti ?

Problème n°95bis (proposé par Loïc Terrier)

On lance un spaghetti de longueur d sur un sol carrelé, les carreaux étant des carrés de côté unité. Quelle est l'espérance de la distance, X_d , entre les centres des deux carreaux où se trouvent les extrémités du spaghetti?

Solution par Jacques Choné (Clermont-Ferrand)

Soit A l'extrémité gauche du spaghetti (on peut négliger le cas où il est vertical, cas de probabilité nulle) et C le carré contenant A (on peut négliger le cas où A est sur la frontière d'un carré, cas de probabilité nulle). On choisit comme origine le centre de C . On suppose que les variables aléatoires x, y, θ sont indépendantes et que x et y suivent la loi uniforme sur $\left[-\frac{1}{2}, \frac{1}{2}\right]$ et θ la loi uniforme sur $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. La seconde extrémité, B , du spaghetti a alors pour coordonnées $(x + d \cos \theta, y + d \sin \theta)$. On a, en désignant par $\text{round}(\alpha)$ l'entier le plus proche de α :

$$X_d = \sqrt{(\text{round}(x + d \cos \theta))^2 + (\text{round}(y + d \sin \theta))^2}$$

$$E(X_d) = \frac{1}{\pi} \int_{x=-\frac{1}{2}}^{\frac{1}{2}} \int_{y=-\frac{1}{2}}^{\frac{1}{2}} \int_{\theta=-\frac{\pi}{2}}^{\frac{\pi}{2}} X_d d\theta dy dx .$$

Pour des raisons de symétrie (ou par des petits changements de variables appropriés) on a aussi:

$$E(X_d) = \frac{4}{\pi} \int_{x=-\frac{1}{2}}^{\frac{1}{2}} \int_{y=-\frac{1}{2}}^{\frac{1}{2}} \int_{\theta=0}^{\frac{\pi}{4}} X_d d\theta dy dx .$$

Il semble difficile d'en tirer une expression "simple" en fonction de d . Toutefois on peut calculer des approximations de $E(X_d)$, par exemple, par simulation de Monte-Carlo (penser aussi à la loi des grands nombres appliquée à n variables aléatoires indépendantes suivant la même loi que X_d). Voir ci-dessous un programme Maple.

```
>f:=d->(x,y,theta)->sqrt((round(x+d*cos(theta)))^2+(round(y+d*sin(theta)))^2):
>mmc:=proc(n,d) local i,s,x,y,theta;s:=0; for i from 1 to n do
x:=stats[random,uniform[-1/2,1/2]](1);
y:=stats[random,uniform[-1/2,1/2]](1);
theta:=stats[random,uniform[0,Pi/4]](1);
s:=s+evalf(f(d)(x,y,theta)) od;evalf(s/n) end;
```

5.2.3.Problème 95-bis (suite)

Problème n°95bis-suite (proposé par Christophe Terrier)

Rapport entre la somme des déplacements (longueurs des spaghettis) et la somme des distances entre les sommes de carreaux.

On veut étudier, suivant la valeur de la maille, m , $E(Z)$, avec $Z = \frac{X_{D,m}}{D}$ où

- D , longueur du spaghetti (i.e. du déplacement), suit la loi uniforme sur $[a, b]$ avec $0 < a < b$;
- $X_{D,m} = m \sqrt{\left(\text{round}\left(\frac{x + D \cos \theta}{m}\right)\right)^2 + \left(\text{round}\left(\frac{y + D \sin \theta}{m}\right)\right)^2}$ est la distance des centres des mailles de départ et d'arrivée du déplacement;
- x et y suivent la loi uniforme sur $\left[-\frac{m}{2}, \frac{m}{2}\right]$ et θ suit la loi uniforme sur $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$;
- D, x, y, θ sont indépendantes.

La fonction à étudier est donc la fonction:

$$m \mapsto E(Z) = \frac{1}{b-a} \frac{1}{m} \frac{1}{\pi} \int_{x=-m/2}^{x=m/2} \int_{y=-m/2}^{y=m/2} \int_{\theta=-\pi/2}^{\theta=\pi/2} \int_{D=a}^{D=b} \frac{X_{D,m}}{D} dD d\theta dy dx ,$$

où m prend ses valeurs dans un ensemble "raisonnable" par rapport à $[a, b]$.