



UNIVERSITÉ
DE GENÈVE

FACULTÉ DES SCIENCES
ÉCONOMIQUES ET SOCIALES



Département des Systèmes d'Information

Projet de recherche

Baccalauréat

Analyse et implémentation d'une application prédisant la géolocalité d'une personne

Alexandre Kuhn

Professeur : Dimitri Konstantas

Assistants : Katarzyna Wac
A. Villalba Castro

Août 2008

Table des matières

1	Introduction et but.....	3
2	Recherches effectuées antérieurement.....	5
3	Structure du projet et séquence de fichiers.....	6
3.1	Structure du projet.....	6
3.2	Séquence des fichiers utilisés lors de l'exécution du logiciel.....	8
4	Définition d'une matrice cartographique.....	9
4.1	Transformation des données GPS brutes en matrice.....	9
4.2	Correction d'erreurs des données GPS brutes.....	13
4.3	Création d'une matrice.....	14
5	Algorithme et analyse des données.....	15
5.1	Les données.....	15
5.2	La problématique des valeurs nulles.....	15
5.3	Quel algorithme choisir ?.....	17
5.3.1	Les valeurs nulles.....	17
5.3.2	Justification de prendre en compte les jours de la semaine....	22
6	Implémentation de l'algorithme de prédiction.....	24
6.1	Implémentation sans tenir compte des jours de la semaine.....	24
6.2	Implémentation en tenant compte des jours de la semaine.....	25
7	Evaluation et cas d'utilisation du programme de prédiction.....	26
7.1	Evaluation du programme de prédiction.....	26
7.2	Cas d'étude entre une résolution faible et une résolution élevée..	29
8	Domaine d'utilisation.....	33
9	Conclusion.....	35
9.1	Généralités et objectif atteint.....	35
9.2	Points faibles.....	35
9.3	Futur travail.....	35
	Remerciements.....	37
	Bibliographie.....	38

Chapitre 1

Introduction et but

Aujourd'hui, de plus en plus d'appareils mobiles équipés de GPS sont vendus dans le monde, offrant ainsi aux développeurs de nouveaux horizons à créer des applications novatrices. Le dernier exemple en date, est le i-phone d'Apple.

Selon une étude menée par « ABI Research » [ELBS01], en 2011 il y aura 315 millions d'abonnés au service GPS, contrairement à seulement 12 millions en 2006.

Force est de constater, que d'ici quelques années, la technologie GPS sera totalement démocratisée et qu'elle risque d'être intégrée dans une très grande majorité d'appareils mobiles.

Parallèlement à ce fait, diverses études sociologiques comme par exemple celle du docteur Marta Gonzalez de l'université de Boston (USA) et celle du docteur Webb directeur R&D de l'Ofcom en Angleterre [MPEH01] concluent que l'être humain est un spécimen avec des habitudes extrêmement cadrées. Une personne a une forte tendance à visiter plusieurs fois les mêmes endroits selon le même schéma temporel. D'ailleurs, 80% du temps les gens restent dans un rayon de 30 kilomètres, et dans 50% des cas dans un rayon de 10 kilomètres.

Ces deux constats, laissent percevoir l'éventualité de voir émerger des logiciels capables de prédire le lieu où se trouvera une personne dans un futur proche. Autrement dit, la possibilité de connaître le lieu où se trouvera une personne « X » dans « Y » jours à une heure précise. Pour se faire, il suffit d'enregistrer un historique de coordonnées GPS d'une personne pour déduire les probabilités d'un futur déplacement.

Déjà aujourd'hui, une multitude d'entreprises, que ce soit dans le domaine de la recherche ou le domaine économique, se montrent fortement intéressés par cette technologie. Ceci sera mentionné plus en détails lors du chapitre 8. Mais les principaux secteurs qui s'y intéressent, sont les télécoms, l'assistance médicale mobile (ou « mobil health care » domaine de recherche de l'assistante Katarzyna Wac) ou encore les développeurs d'applications sociales.

La tâche qui m'est octroyée pour mon projet de recherche à l'Université de Genève au sein du groupe de recherche Advanced Systems Group, est de créer une première ébauche d'un programme qui est justement capable de prédire les futurs lieux où se trouvera une personne à une date et heure précise grâce à un historique de données GPS.

Pour conduire cette recherche, je vais me déplacer pendant plus d'un mois avec un GPS, afin d'enregistrer les coordonnées des lieux où je me trouve. Les coordonnées reçues par GPS, sont transmises via Bluetooth à un appareil mobile Qtek 9090 tournant sur l'OS Windows Mobile 2003 pour être transmis via USB sur PC (voir fig.1).

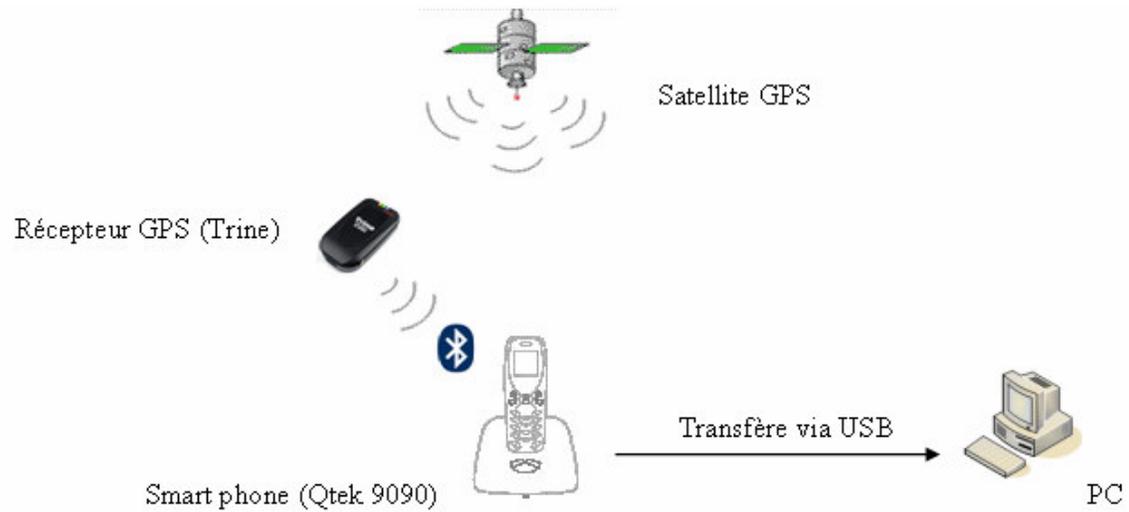


Fig. 1 : Schéma de transmission de données

Une fois les données transmises au PC, elles seront traitées par le programme de prédiction. Le langage de programmation utilisé est Java.

Chapitre 2

Recherches effectuées antérieurement

Avant de commencer la programmation du logiciel de prédiction, diverses revues scientifiques traitant du même sujet ont été étudiées, afin de se familiariser avec l'énoncé.

Un défi à relever pour la prochaine génération de technologie mobil (4G) [USMP01], consiste à fournir aux utilisateurs des services de vidéoconférence ou de vidéo. En sachant que ces services nécessitent un débit considérable, les auteurs de ces thèses cherchent à rendre les réseaux intelligents afin qu'ils réagissent de façon proactive. Concernant les réseaux utilisant l'ATM [MMLC01] et le Wireless [MPMA01], [MPWN01], c'est le même cas de figure. Ce sont des technologies qui sont de plus en plus utilisées par le grand public. Cependant, les recherches effectuées jusqu'à aujourd'hui concernant la prédiction de localité pour assurer une meilleure garantie de la « *quality-of-service* » (QoS) sont restées faibles. Du fait que l'ATM établit une connectivité avant de faire le transfert des paquets, il serait d'un énorme bénéfice de pouvoir prédire le futur positionnement géographique d'un utilisateur. Il faudrait être capable d'anticiper les déplacements d'un mobil afin de réserver les ressources nécessaires pour la garantie de la QoS. Cela, dans le but de ne pas surcharger inutilement le réseau en demandant à un grand nombre de nœuds une future connexion qui n'aura finalement pas lieu. Il s'agit d'algorithmes capables de prédire les micros déplacements.

Il n'est nulle part fait allusion aux algorithmes de prédiction prenant en considération l'historique de positionnement dans ces travaux. La plupart de ces thèses se basent sur l'implémentation de logiciel pouvant déterminer de façon relative [RPCD01] la prochaine cellule parcourue par l'agent mobil. Ce sont donc des cas d'étude qui effectuent la prédiction grâce au réseau GSM [MDET01]. En effet, les chercheurs de ces thèses utilisent le modèle de Markov [ULSL01] pour la prédiction. Mais, le problème avec cette méthode, c'est qu'uniquement le lieu suivant peut être prédit. Une prédiction dans un futur plus lointain n'est pas possible.

Cela dit, les auteurs n'utilisent pas la technologie GPS. Ils justifient leur choix par le fait que c'est trop coûteux. Ces travaux ont été réalisés il y a plus de cinq ans et les choses ont évolué depuis. Aujourd'hui, l'utilisation du GPS a considérablement augmenté et c'est une tendance qui va continuer.

Toutefois, une étude menée par Microsoft [PWWT01], [IDPT01] parle du bien-fondé de l'utilisation de la technologie GPS. L'avantage principal du GPS, c'est qu'il n'a pas besoin d'infrastructure cellulaire comme par exemple le GSM. La problématique des zones d'ombres peut être facilement résolue en utilisant des algorithmes intelligents. Même si le signal du GPS ne passe pas dans un bâtiment, le système est capable de connaître la localité de l'utilisateur grâce aux dernières données de GPS enregistrées.

Chapitre 3

Structure du projet et séquence de fichiers

3.1 *Structure du projet*

Grâce à des données GPS collectées, le but de ce projet sera de prédire les futurs lieux où une personne se trouvera à une date et une heure précises. Ce travail se divise en trois phases principales (voir fig. 2). C'est lors des trois chapitres suivants, que ces phases seront traitées en détail. En voici un aperçu :

- 1) La première phase consiste à créer une matrice de coordonnées géographiques relatives d'après des coordonnées GPS (dans notre cas, c'est une carte de Genève qui est utilisée). Ces données auront une signification pour l'utilisateur. Cependant, les coordonnées GPS brutes ne sont pas faciles à interpréter.
- 2) Suivront divers tests nécessaires à la conception de l'algorithme. Ces tests seront effectués avec le logiciel Matlab. Ce programme présente l'avantage de pouvoir manipuler des matrices avec une grande facilité.
- 3) Finalement, une fois la bonne méthodologie trouvée, il restera à implémenter l'algorithme en Java. C'est lors de ce point que la prédiction se fait.

Structure du projet

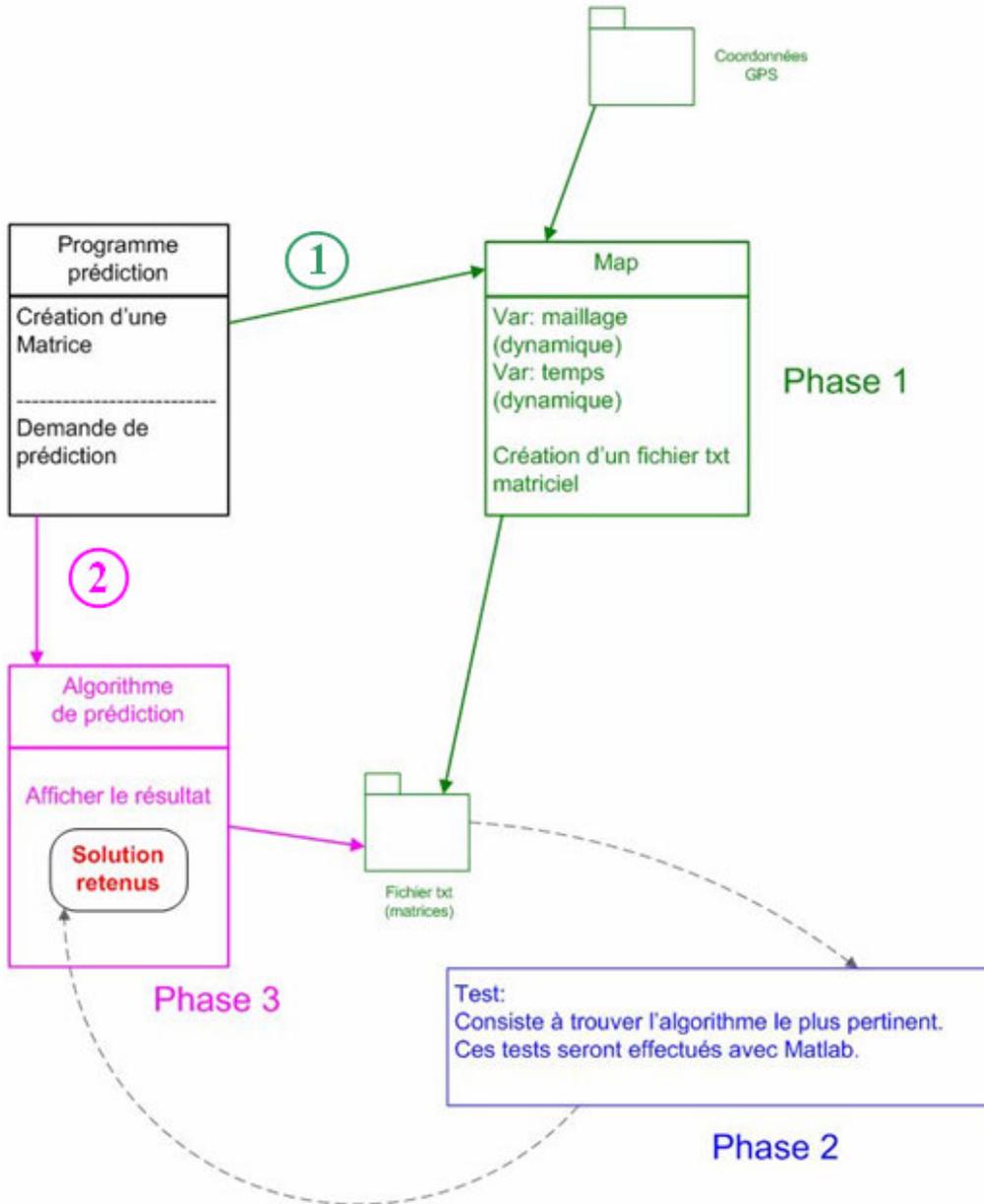


Fig. 2 : Les trois phases du projet

3.2 Séquence des fichiers utilisés lors de l'exécution du logiciel

Comme il en sera question dans le chapitre suivant, les données GPS envoyées au Qtek ne sont pas exemptes d'erreurs. C'est un fait qui doit être pris sérieusement en compte lors de la conception du programme. Par conséquent, un fichier intermédiaire (en .txt), très semblable au fichier obtenu par le Qtek (en .csv) sera créé, qui ne présente aucune erreur. Grâce à ce fichier débogué un fichier (également en .txt) matriciel sera généré. Enfin, c'est grâce à ce fichier matriciel que la prédiction sera calculée (voir fig. 3).

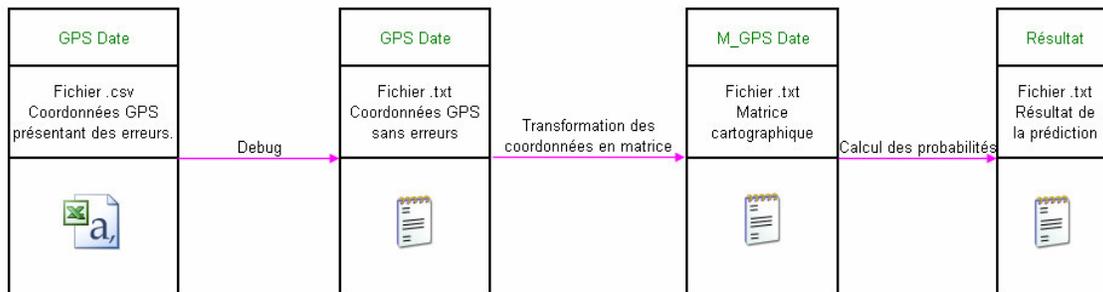


Fig. 3 : Séquence des fichiers utilisés

Chapitre 4

Définition d'une matrice cartographique

4.1 Transformation des données GPS brutes en matrice

Les données GPS sont récupérées en format .csv du Qtek sous formes de texte brut (voir fig. 4). L'application utilisée pour enregistrer ces données est GVT.exe. Elle a été créée par Pravin Pawar de l'Université Twente au Pays-Bas.

```
DATE, TIME, LAT, LONG, SPEED, HEADING  
2008-06-17,10:28:56,46.2000233333333334,6.14213,0.564521084,310.63  
2008-06-17,10:28:57,46.2000216666666665,6.14213,0.564687764,310.63
```

Fig. 4 : données GPS (enregistrement de deux secondes)

Explication des données GPS :

Elles sont séparées par une virgule. Les six indications suivantes sont présentes dans la chaîne de caractère :

- La date : elle est formatée selon les normes américaines. Soit AAAA-MM-JJ
- L'heure : elle est exprimée en heure GMT. Le format utilisé est classique. Soit HH:MM:SS
- La latitude : expression de la position d'un point sur Terre (du nord ou au sud de l'équateur)
- La longitude : expression du positionnement est-ouest d'un point sur Terre (la référence est le méridien de Greenwich).
- La vitesse : cette valeur indique la vitesse de déplacement lorsque l'utilisateur est en mouvement. Cette donnée ne sera pas utilisée lors de ce projet.
- Le cap : indication de la direction lors d'un déplacement de l'utilisateur. Cette donnée ne sera pas utilisée lors de ce projet.

Erreurs générées par le logiciel GPS ou le Qtek 9090 :

Comme mentionné lors du premier chapitre, ces données sont extraites directement d'un mobil Qtek 9090. L'enregistrement des points GPS se fait avec un intervalle d'une seconde (voir fig. 4). La durée d'enregistrement ne joue en principe aucun rôle, mais il est conseillé de faire une sauvegarde toutes les 24 heures, afin d'éviter de perdre de longues plages de coordonnées. L'envoi des données du GPS mobil Trine au Qtek, peut s'avérer défaillant par moment. Raison pour laquelle, lorsqu'il y a un dysfonctionnement

de transmission, il est conseillé de faire un redémarrage des deux appareils en question. Comme il est expliqué au chapitre 4.2, le programme peut gérer ce type de défaut, mais va inévitablement perdre en précision lors de la prédiction. Les pertes de ces données peuvent, par exemple survenir lorsque l'accu du Qtek est déchargé. Il est particulièrement conseillé de surveiller le niveau de batterie.

Une fois le fichier acquis, il va être lu par le programme et une matrice sera générée.

Le but de cette première étape :

Il s'agit de générer dynamiquement une matrice variable à partir des données GPS brutes. La matrice doit être capable de représenter une localité avec un maillage et un intervalle de temps défini par l'utilisateur.

Que représente cette matrice ?

Les valeurs sont des références spatiales type cartographiques ou des lieux qui sont définis par l'utilisateur. Ceux-ci se font automatiquement lorsque la matrice est générée pour la première fois. Comme pour une carte routière, il y a un découpage nord – sud et est – ouest. La valeur est attribué selon le désir de l'utilisateur. En voici 3 exemples :

1) Découpage cartographique (Exemple : I4)

C'est exactement comme pour une carte dans le sens classique. Ce sont des données relatives (voir fig. 5-P2).

2) Découpage cartographique en numérique (Exemple : 0.0904)

09 (09 = I) représente la latitude alors que 04 (04 = 4) désigne la longitude (voir fig. 5-P2). C'est le même principe que pour le point précédent, mais les données sont uniquement représentées en numérique. C'est d'une grande utilité lors de calculs matriciels (comme par exemple avec Matlab).

3) Renommage selon les préférences de l'utilisateur (Exemple : Unimail)

L'utilisateur peut renommer les valeurs pour leur attribuer une signification (voir fig. 5-P1).

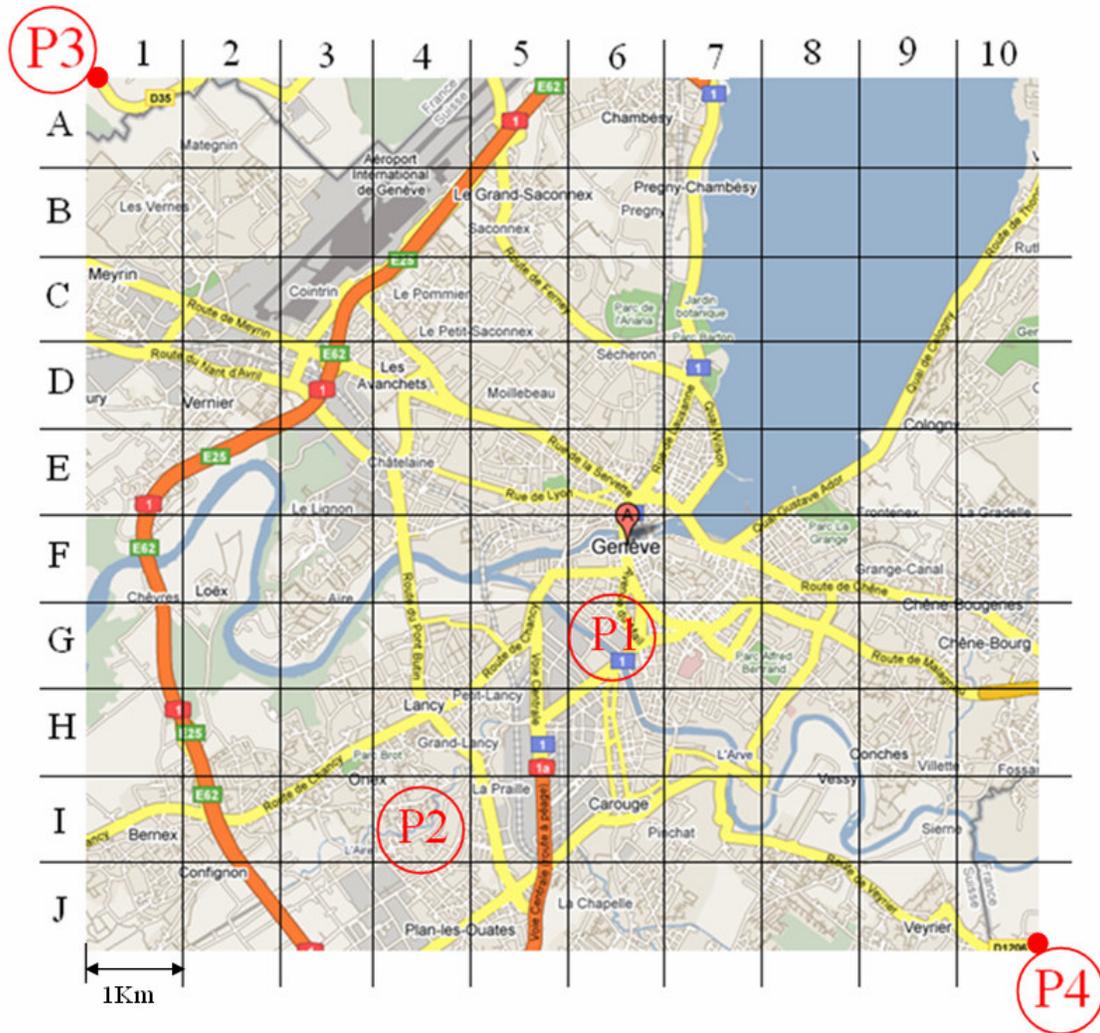


Fig. 5 : Carte de Genève (Google Maps)

Lors de cet exemple, c’est un maillage de 10 x 10 qui est appliqué. Cette valeur peut être modifiée selon le désir de l’utilisateur à tout moment. Si l’utilisateur fait le choix de prendre, par exemple, un maillage de 20 x 20, alors la position de P2 passe de I4 à S8 ou T9. La précision est donc multipliée par deux.

Structure de la matrice :

Maintenant qu’on a vu la signification des valeurs de la matrice, il ne reste plus qu’à comprendre sa structure. Comme on l’a dit, il s’agit d’un fichier .txt. Les valeurs y sont séparées par un espace blanc.

Une ligne horizontale de la matrice constitue un jour, alors que les lignes verticales représentent les valeurs espacées par un laps de temps défini par l’utilisateur. Autrement dit, plus on demande un échantillonnage élevé, plus la ligne sera longue. Par exemple, s’il y a extraction de données une fois par heure, alors une ligne (donc un jour) aura une taille

4.2 Correction d'erreurs des données GPS brutes

Les fichiers GPS extrait du Qtek, ne sont malheureusement pas exempts d'erreurs. Si ces défauts ne sont pas corrigés dès le début, la matrice peut être incorrecte et des erreurs d'exécution peuvent survenir lors de la lecture du programme. Pour faciliter le travail de la création de la matrice et de la prédiction, un fichier intermédiaire sans erreurs est généré.

Voici des cas d'erreurs qui peuvent survenir lors de la récupération de données GPS et les corrections apportées :

Sources d'erreurs

```
2008-06-19,16:33:30,46.19974833,6.142351666,0.72,196.33
2008-06-19,16:33:31,46.19974,6.14235,0.6802,196.33
DATE, TIME, LAT, LONG, SPEED, HEADING
2008-06-19,16:33:32,46.199733333,6.1423483333,0.74,196.33
2008-06-19,16:33:33,46.199726666,6.14235,0.800,196.33
2008-06-19,16:33:34,46.199723333,6.142353333,0.7403,196.33
```

Corrections

```
2008-06-19,16:33:30,46.19974833,6.142351666,0.72,196.33
2008-06-19,16:33:31,46.19974,6.14235,0.6802,196.33
2008-06-19,16:33:32,46.199733333,6.1423483333,0.74,196.33
2008-06-19,16:33:33,46.199726666,6.14235,0.800,196.33
2008-06-19,16:33:34,46.199723333,6.142353333,0.7403,196.33
```

Fig. 7 : Présence de chaînes de caractères

```
2008-06-19,09:06:43,46.19990166,6.1421616666,0.182088,346.47
2008-06-19,09:06:44,46.1999,6.14216,0.2499107,346.47
2008-06-19,09:06:45,46.19989666,6.142158333,0.3195,346.47
2008-06-19,09:06:46,46.19989666,6.142156666,0.125,346.47
2008-06-19,09:06:47,46.199895,6.142155,0.153423384,346.47
```

```
2008-06-19,09:06:43,46.19990166,6.1421616666,0.182088,346.47
2008-06-19,09:06:44,46.1999,6.14216,0.2499107,346.47
2008-06-19,09:06:45,46.19989666,6.142158333,0.3195,346.47
2008-06-19,09:06:46,46.19989666,6.142156666,0.125,346.47
2008-06-19,09:06:47,46.199895,6.142155,0.153423384,346.47
```

Fig. 8 : Lignes sans données

```
2008-06-19,17:11:19,46.199715,6.142466666,0.36802,196.33
2008-06-19,17:11:20,46.199715,6.14,0.337343652,196.33
2008-06-19,17:11:21,46.199715,6.1424633,0.30,196.33
2008-06-19,17:11:21,46.199715,6.1424633,0.30,196.33
2008-06-19,17:11:22,46.199715,6.142461666,0.35,196.33
2008-06-19,17:11:23,46.199715,6.14246,0.298,196.33
```

```
2008-06-19,17:11:19,46.199715,6.142466666,0.36802,196.33
2008-06-19,17:11:20,46.199715,6.14,0.337343652,196.33
2008-06-19,17:11:21,46.199715,6.1424633,0.30,196.33
2008-06-19,17:11:22,46.199715,6.142461666,0.35,196.33
2008-06-19,17:11:23,46.199715,6.14246,0.298,196.33
```

Fig. 9 : Répétition de la même seconde

```
2008-06-20,03:57:15,46.198935,6.142,0.0,147.66
2008-06-20,03:57:16,46.19883,6.142338,0.0,147.66
2008-06-20,03:57:17,46.19876,6.142375,0.0,147.66
2008-06-20,03:57:19,46.1986833,6.142386,0.0,147.66
2008-06-20,03:57:20,46.198671666,6.14239,0.0,147.66
2008-06-20,03:57:21,46.19862166,6.142388,0.0,147.66
```

```
2008-06-20,03:57:15,46.198935,6.142,0.0,147.66
2008-06-20,03:57:16,46.19883,6.142338,0.0,147.66
2008-06-20,03:57:17,46.19876,6.142375,0.0,147.66
2008-06-20,03:57:18,0,0,0
2008-06-20,03:57:19,46.1986833,6.142386,0.0,147.66
2008-06-20,03:57:20,46.198671666,6.14239,0.0,147.66
2008-06-20,03:57:21,46.19862166,6.142388,0.0,147.66
```

Fig. 10 : Saut dans le temps (perte de données)

Le programme corrige les erreurs suivantes :

- Figure 7 : Des chaînes de caractères sont supprimées.
- Figure 8 : Des lignes vides sont supprimées.
- Figure 9 : Il y a des doublons. Dans de rares cas, le Qtek enregistre deux fois par seconde des coordonnées en double. Elles sont supprimées.

- Figure 10 : Des trous de 1 ... n secondes peuvent être présents. Du fait que ce sont des valeurs inconnues, la valeur 0 est attribuée pour la latitude et la longitude. C'est lors de la prédiction que ces valeurs seront interprétées de façon plus subtile.
C'est sans doute la correction la plus fréquente, car elle permet de remplir de longues plages de temps en cas d'oubli de l'utilisateur d'enregistrer les coordonnées.

4.3 Création d'une matrice

La première étape du programme définit la taille de la matrice. Il faut que cette matrice s'adapte automatiquement aux nombres de jours qui sont présents dans le fichier GPS. Il faut qu'elle prenne en considération l'heure de départ de ce fichier pour synchroniser le départ à l'endroit approprié de la matrice. En revanche, définir l'intervalle de temps est beaucoup plus simple. C'est l'utilisateur qui le définit directement lors de l'exécution du programme.

Lors de la deuxième étape, le programme remplit la matrice avec des valeurs cartographiques. Comme expliqué au chapitre 4.1, ces valeurs peuvent être définies sous diverses formes. Le programme nécessite que l'utilisateur lui définisse un point de départ tout comme un point d'arrivée exprimée en latitude et longitude. Finalement, il ne reste à l'utilisateur plus qu'à prendre la décision du maillage désiré. En fonction de l'intervalle de temps choisi, le programme va parcourir le fichier GPS et extraire les données à l'heure appropriée. Autrement dit, si l'utilisateur choisit par exemple une précision d'une minute, seule une ligne sur 60 sera extraite (l'échantillonnage est donc de 60:1). Mais cela seulement après que la synchronisation de l'heure de départ a eu lieu.

La latitude et la longitude de la ligne extraite, sont transformées en positions cartographiques selon le désir de l'utilisateur comme on l'a vu dans le chapitre 4.1. Grâce aux points de départs définis, les points GPS absolus sont convertis en valeurs relatives. Lors de ce projet, mis à part les analyses effectuées avec Matlab, le positionnement est exprimé en découpage cartographique (exemple I4). Conformément au maillage indiqué par l'utilisateur, un point cartographique est ainsi généré. Cette opération se répète jusqu'à ce que la fin du fichier GPS soit atteinte. Finalement un fichier .txt de la matrice est enregistré.

Chapitre 5

Algorithme et analyse des données

5.1 Les données

Les coordonnées GPS utilisées pour faire les tests sont celles de Katarzyna Wac. Ces données ont été enregistrées pendant une période d'un mois (du samedi 17 novembre 2007 – samedi 15 décembre 2007). Cependant, seuls les points de départs et d'arrivées sont présents. Ces données ont été enregistrées à la main dans un carnet. Il a donc fallu créer manuellement un fichier .csv pour insérer toutes ces données collectées pendant le mois en question. Grâce à Excel, une incrémentation du temps par seconde est facilement réalisable. Une fois le calendrier d'un mois établi, les coordonnées GPS sont insérées dans le tableau. Concernant la variation des coordonnées pendant les déplacements, elles sont incrémentées ligne par ligne en respectant l'intervalle de temps, induisant ainsi en un résultat de déplacement « artificiel » qui est linéaire. Il est évident que cette méthode n'est pas représentative de la réalité. Mais c'est la seule manière de surmonter la problématique de la non-couverture temporelle de l'enregistrement. Par ailleurs, la différence entre les données enregistrées artificiellement et celles de la vie réelle s'avère être minime.

Finalement, prenons note qu'il y a de longues plages où aucun positionnement n'est donné. Il y a notamment la présence de trois jours (le 23, 24, 25 novembre) qui ont des valeurs en blanc. C'est lors du sous-chapitre suivant qu'on va s'attarder sur cette problématique.

Tout comme l'assistante Katarzyna Wac, j'enregistre moi-même mes déplacements. Malheureusement, la période de recueil s'avère être actuellement trop courte. Arrivé à ce stade du projet, il est nécessaire d'entamer la phase d'analyse pour trouver le bon algorithme de prédiction. Comme j'ai seulement une semaine de données à ma disposition (du 24 juillet 2008 au 31 juillet 2008), la masse de celles-ci s'avère être beaucoup trop faible. C'est pour cette raison que je vais me baser principalement sur les données de Katarzyna Wac lors de ce chapitre.

5.2 La problématique des valeurs nulles

C'est avec le logiciel Matlab que les données matricielles sont interprétées et analysées. Comme mentionné précédemment, il y a des tranches d'horaires où aucune donnée n'est présente (voir fig. 11). Ces données nulles, sont présentes dans tous les fichiers, que ce soit dans celui des coordonnées GPS contenant des erreurs (.csv), dans le fichier débogué

(.txt) ou dans le fichier matrice (txt). C'est le programme de prédiction qui va interpréter ces valeurs nulles (voir chapitre 5.3.1).

```
2007-11-17,22:53:54,0.0,0.0,0.0,0.0
2007-11-17,22:53:55,0.0,0.0,0.0,0.0
2007-11-17,22:53:56,0.0,0.0,0.0,0.0
2007-11-17,22:53:57,0.0,0.0,0.0,0.0
2007-11-17,22:53:58,0.0,0.0,0.0,0.0
2007-11-17,22:53:59,0.0,0.0,0.0,0.0
2007-11-17,22:54:00,0.0,0.0,0.0,0.0
2007-11-17,22:54:01,0.0,0.0,0.0,0.0
2007-11-17,22:54:02,0.0,0.0,0.0,0.0
2007-11-17,22:54:03,0.0,0.0,0.0,0.0
2007-11-17,22:54:04,0.0,0.0,0.0,0.0
2007-11-17,22:54:05,0.0,0.0,0.0,0.0
2007-11-17,22:54:06,0.0,0.0,0.0,0.0
2007-11-17,22:54:07,0.0,0.0,0.0,0.0
2007-11-17,22:54:08,0.0,0.0,0.0,0.0
2007-11-17,22:54:09,0.0,0.0,0.0,0.0
2007-11-17,22:54:10,0.0,0.0,0.0,0.0
2007-11-17,22:54:11,0.0,0.0,0.0,0.0
2007-11-17,22:54:12,0.0,0.0,0.0,0.0
2007-11-17,22:54:13,0.0,0.0,0.0,0.0
2007-11-17,22:54:14,0.0,0.0,0.0,0.0
2007-11-17,22:54:15,0.0,0.0,0.0,0.0
2007-11-17,22:54:16,0.0,0.0,0.0,0.0
2007-11-17,22:54:17,0.0,0.0,0.0,0.0
2007-11-17,22:54:18,0.0,0.0,0.0,0.0
2007-11-17,22:54:19,0.0,0.0,0.0,0.0
2007-11-17,22:54:20,0.0,0.0,0.0,0.0
2007-11-17,22:54:21,0.0,0.0,0.0,0.0
2007-11-17,22:54:22,0.0,0.0,0.0,0.0
2007-11-17,22:54:23,0.0,0.0,0.0,0.0
2007-11-17,22:54:24,0.0,0.0,0.0,0.0
2007-11-17,22:54:25,0.0,0.0,0.0,0.0
```

Fig. 11 : Données GPS contenant des valeurs zéro

En effet, l'importance de ces données nulles est de **29.3%** pour Katarzyna Wac.

A titre comparatif, un deuxième test est effectué sur mes données. La période d'évaluation porte seulement sur une semaine. Le résultat est de **33.88%** de pertes de données.

Ces résultats sont quelque peu inquiétants et laissent présumer une perte de précision lors de la prédiction.

L'explication de ces plages vides est principalement due aux deux facteurs suivants :

- **L'erreur humaine**

Pour éviter les plages nulles, il est évident que l'utilisateur est tenu de surveiller rigoureusement l'enregistrement de ses données. Il est particulièrement important de s'assurer de la bonne santé de l'accu. A noter cependant, que le Qtek n'est pas fiable à 100% et qu'il peut se bloquer selon certaines circonstances.

- **Les zones d'ombres**

C'est un problème largement connu par les utilisateurs de GPS. Ces zones d'ombres sont particulièrement présentes en ville, car la réception d'un ou

plusieurs satellites peu être gênée par les bâtiments environnants, rendant ainsi la triangulation impossible.

Par ailleurs, les probabilités de pertes de données sont considérablement plus élevées lorsque l'utilisateur se trouve à l'intérieur d'un bâtiment. C'est la raison la plus fréquente de pertes de données.

Un deuxième modèle de prédiction devient alors concevable et sera validé lors du chapitre 5.3. Il consiste à corriger les plages de valeurs nulles en les remplaçant par les dernières coordonnées disponibles. Ce modèle se justifie principalement par le fait, qu'une personne entrant dans un lieu non couvert par le réseau GPS est supposée y rester jusqu'à ce que de nouvelles coordonnées GPS soient réceptionnées.

5.3 Quel algorithme choisir ?

5.3.1 Les valeurs nulles

Comme on vient de constater, deux méthodes de prédictions sont possibles (voir fig. 12). La base de ces deux algorithmes est la même. En effet, c'est le poids des diverses coordonnées sur l'ensemble de l'heure des journées enregistrées qui vont donner un pourcentage de probabilité pour chaque localité.

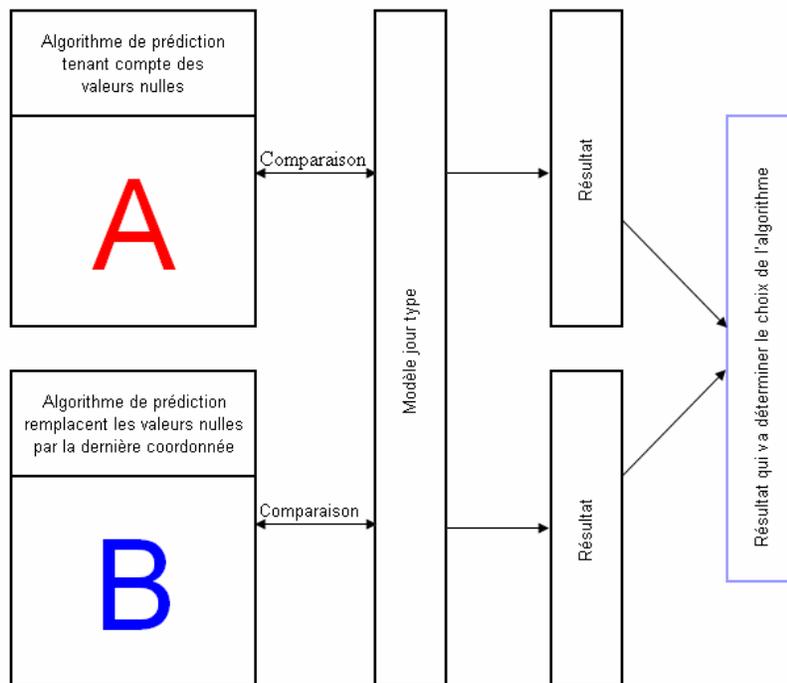


Fig. 12 : Séquence de comparaison des deux algorithmes

La matrice utilisée lors de l'essai est composée d'un intervalle de temps de cinq minutes avec un maillage de 10 x 10.

Entre 00:00 et 24:00 heures ce sont donc 300 tirages qui sont effectués.

Le premier algorithme affichera le pourcentage des valeurs nulles (voir fig. 12 lettre « A »), alors que le deuxième algorithme charge les valeurs nulles avec la dernière donnée disponible (voir fig. 12 lettre « B »).

Justification du choix des cinq minutes :

Idéalement, il aurait fallu faire le test pour chaque seconde. Cependant, à chaque fois que la précision du temps est doublée, les intervalles de la matrice doivent être parcourus deux fois plus souvent avec une précision également deux fois supérieure (utilisation d'une matrice deux fois plus grande). Le temps de calcul est donc exponentiel. Même en prenant une minute d'intervalle, le temps que nécessite la machine à faire les calculs est exorbitant. Par conséquent, l'intervalle de cinq minutes est un compromis idéal.

Ces deux méthodes sont comparées à un modèle d'une journée type de l'assistante Katarzyna Wac, afin d'analyser la pertinence de celles-ci.

Le modèle d'une journée type est une séquence de lieux créés artificiellement qui reflète au mieux l'emplacement de la personne. Elle est simplificatrice afin de permettre un maximum de corrélation entre l'ensemble des journées et le modèle. Par ailleurs, Les phases de déplacement sont volontairement supprimées, la aussi, dans le souci d'augmenter au maximum les correspondances.

Composition du modèle d'une journée type

Les informations sont extraites des données réelles. C'est-à-dire, elles sont reprises du carnet de positionnement de Katarzyna Wac. Comme il a été mentionné lors du début de ce chapitre, les données de l'assistante ont été enregistrées à la main. Dans un souci de simplifier au maximum cette analyse, seules les heures passées à domicile et à l'université sont employées. Par conséquent, le tableau suivant (fig. 13) démontre l'heure de basculement entre le départ du domicile et l'arrivée au bureau. Quand au deuxième tableau (fig. 14), il illustre l'heure de basculement du départ du bureau à l'arrivée à domicile.

Par ailleurs, les valeurs nulles ne sont pas prises en compte.

Dates	L'heure d'arrivée au travail	Balance du nb d'heures	
		Les heures tôt	Les heures tard
17-Nov-07	No Data		
18-Nov-07	11:20		11:20
19-Nov-07	7:50	7:50	
20-Nov-07	9:33		9:33
21-Nov-07	8:10	8:10	
22-Nov-07	8:30	8:30	
23-Nov-07	5:45	5:45	
24-Nov-07	No Data		
25-Nov-07	No Data		
26-Nov-07	9:02		9:02
27-Nov-07	8:36	8:36	
28-Nov-07	8:03	8:03	
29-Nov-07	9:08		9:08
30-Nov-07	12:22		12:22
1-Dec-07	No Data		
2-Dec-07	11:43		11:43
3-Dec-07	7:52	7:52	
4-Dec-07	9:19		9:19
5-Dec-07	8:03	8:03	
6-Dec-07	9:51		9:51
7-Dec-07	8:41	8:41	
8-Dec-07	No Data		
9-Dec-07	12:11		12:11
10-Dec-07	9:10		9:10
11-Dec-07	8:45		8:45
12-Dec-07	8:40	8:40	
13-Dec-07	9:20		9:20
14-Dec-07	6:56	6:56	
15-Dec-07	No Data		
Total Jours	23	Total Jours	12

Fig. 13: Tableau illustrant le basculement de l'heure (pour le matin)

Dates	L'heure d'arrivée à domicile	Balance du nb d'heures	
		Les heures tôt	Les heures tard
17-Nov-07	23:23		23:23
18-Nov-07	19:10	19:10	
19-Nov-07	18:00	18:00	
20-Nov-07	18:10	18:10	
21-Nov-07	20:46		20:46
22-Nov-07	19:48	19:48	
23-Nov-07	No Data		
24-Nov-07	No Data		
25-Nov-07	No Data		
26-Nov-07	19:26	19:26	
27-Nov-07	23:12		23:12
28-Nov-07	22:30		22:30
29-Nov-07	No Data		
30-Nov-07	18:50	18:50	
1-Dec-07	No Data		
2-Dec-07	21:41		21:41
3-Dec-07	19:30	19:30	
4-Dec-07	20:04		20:04
5-Dec-07	23:54		23:54
6-Dec-07	18:08	18:08	
7-Dec-07	19:00	19:00	
8-Dec-07	No Data		
9-Dec-07	23:49		23:49
10-Dec-07	19:06	19:06	
11-Dec-07	20:44		20:44
12-Dec-07	18:11	18:11	
13-Dec-07	23:17		23:17
14-Dec-07	22:07		22:07
15-Dec-07	No Data		
Total Jours	22	Total Jours	11

Du au même nb de jour:
 $(20:44 - 19:48) / 2 = 28\text{min}$
 $19:48 + 28\text{ min} = 20:16 \rightarrow \mathbf{20:15}$

Fig. 14: Tableau illustrant le basculement de l'heure (pour le soir)

Cette journée est donc composée de la façon suivante :

00 :00 – 08 :45 → domicile
08 :45 – 20 :15 → travail
20 :15 – 24 :00 → domicile

Explications des deux tableaux :

Ce sont les horaires d'arrivée qui sont représentés. Lors du premier tableau, on retrouve l'ensemble des heures d'arrivée au bureau, respectivement sur celui du bas, toutes les heures d'arrivée au domicile.

Pour retrouver la plage horaire du lieu dominant, les différentes heures des jours disponibles sont divisées en deux groupes. Le premier groupe représente les horaires les plus tôt, alors que le deuxième, les horaires les plus tard. Les deux groupes sont triés afin d'avoir le même nombre de jours d'un groupe à l'autre.

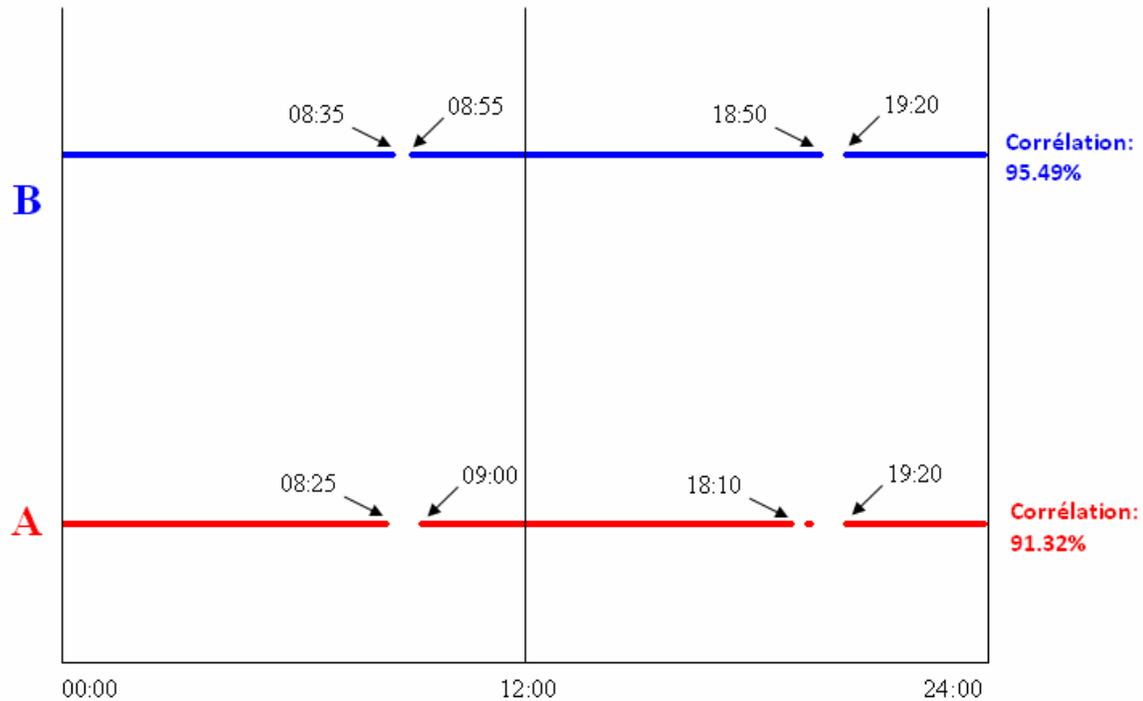
Sur le premier tableau, l'heure de basculement est 08:45 (l'heure la plus tôt des valeurs du groupe tard).

Lors du deuxième tableau, les nombres de jours d'un groupe à l'autre sont identiques. Par conséquent, c'est l'heure intermédiaire entre l'heure la plus tard du groupe d'heures tôt et l'heure le plus tôt du groupe d'heures le plus tard qui est pris comme résultat.

En résumé, c'est toujours la valeur dominante qui est prise en compte.

Résultat de la comparaison

Comme on peut le constater à la figure 15, les résultats sont très encourageants. Les deux algorithmes ont de bonnes performances. Cependant, celui qui prend en compte les dernières valeurs enregistrées est encore bien plus efficace.



Temps allant de 00:00 heure à 24:00 heure (de 0% - 100% de corrélation)
 Test pour 5 minutes. La ligne est donc représentée par 300 points

Fig. 15 : Comparaison de la corrélation avec le modèle de jour type des deux algorithmes

Explication du graphique :

Les trous dans les lignes continues expriment des erreurs de corrélation avec le modèle d'une journée type.

Dans l'idéal, la ligne devrait être continue de 00 :00 – 24 :00 heure. Dans ce cas, on aurait une corrélation de 100%.

Ce graphique présente l'avantage de connaître l'heure à laquelle il n'y a pas de corrélation. Le positionnement géographique est cependant inconnu.

Analyse des périodes sans corrélation :

En se référant au modèle d'une journée type, on constate qu'il y a problème de corrélation lors de l'arrivée au travail respectivement au domicile le soir. Deux raisons expliquent ce fait. Premièrement, les déplacements ne sont pas pris en compte avec le modèle d'une journée type et deuxièmement, l'assistante enregistrerait manuellement l'heure et le lieu d'arrivée. Il y a donc un certain laps de temps pendant lequel aucune donnée n'est enregistrée. Cela n'est évidemment pas le cas pour le modèle d'une journée type.

C'est précisément lors de ce cas de figure où l'algorithme remplace les valeurs nulles par la dernière valeur obtenue surpasse son concurrent.

Ce constat est très positif pour contrer le problème des zones d'ombres, car le cas de figure est exactement le même. Une personne entre dans un bâtiment et la position n'est souvent plus connue. Cependant, les dernières coordonnées le sont et il suffit de les appliquer à l'ensemble du temps où les valeurs nulles sont présentes.

Conclusion : C'est l'algorithme remplaçant les valeurs nulles par les dernières données disponibles qui est choisi pour l'intégration finale.

5.3.2 Justification de prendre en compte les jours de la semaine

Comme constaté, l'algorithme fonctionne bien avec le modèle de jour type. Mais, que se passe-t-il si un jour « X » ne ressemble pas au modèle de jour type ? Il paraît évident, que par exemple les week-ends où une activité sportive à un jour précis, ne sont pas sujets à de fortes corrélations.

Comme précédemment, un test comparatif est effectué avec le modèle de jour type. Un jour de la semaine, le plus **improbable**, y est comparé. Malheureusement, ce test n'est pas d'une grande exactitude, car le nombre de jours enregistrés est de quatre et c'est trop faible. Néanmoins, cela permettra de connaître la portée du phénomène.

L'assistante Katarzyna Wac a passé trois dimanches sur quatre à l'université. Ce jour n'est donc pas approprié pour trouver une corrélation faible. En revanche, elle a passé deux samedis sur trois à la maison (le quatrième samedi étant inconnu).

Le jour le plus improbable à une corrélation est donc le samedi et voici le résultat (fig. 16) :

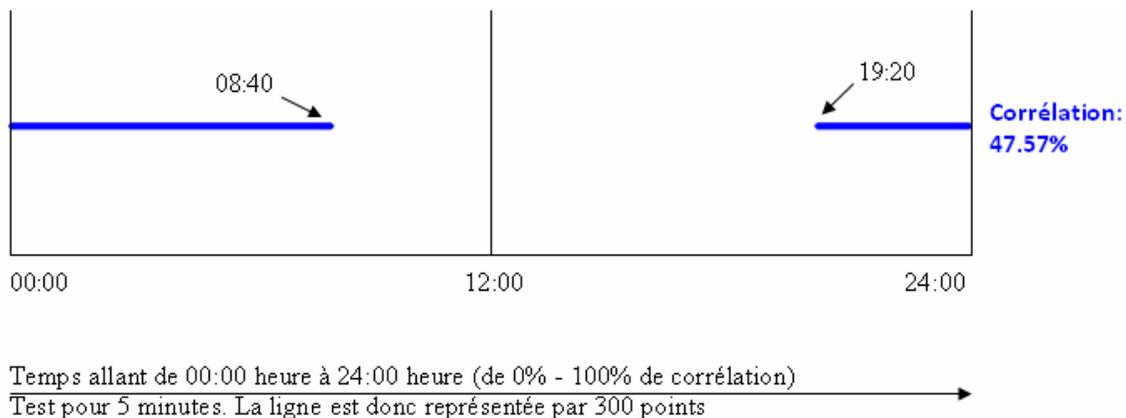


Fig. 16 : corrélation du modèle de jour type avec uniquement le samedi

Dès lors, il paraît évident qu'un groupement des jours de la semaine est nécessaire. Comme décrit plus haut, un enregistrement sur plusieurs mois fait défaut. La décision suivante a été prise : si l'enregistrement des données dépasse trois semaines, alors le programme va tenir compte des jours spécifiques de la semaine. Si la durée

d'enregistrement devait être plus faible que trois semaines, alors la probabilité sera calculée sur la base de tous les jours de la semaine.

Du fait de la courte durée de l'enregistrement, aucun test n'a pu être effectué pour trouver le nombre de jours idéal à passer d'un système de prédiction à l'autre. Attention, la décision des trois semaines est purement **arbitraire** et aucun test scientifique ne prouve que ce soit la bonne décision. Ce point serait définitivement à revoir lors d'un futur travail concernant ce projet.

Chapitre 6

Implémentation de l'algorithme de prédiction

L'implémentation de l'algorithme de prédiction se divise donc en deux parties. La première ne tient pas compte des jours de la semaine au cas où l'enregistrement ne devrait dépasser les trois semaines, alors que la deuxième partie en tient compte.

6.1 Implémentation sans tenir compte des jours de la semaine

C'est quasiment une traduction 1:1 de l'algorithme Matlab en Java.

L'utilisateur est amené à entrer l'heure de la prédiction désirée. Le programme va prendre toutes les valeurs cartographiques correspondant à l'heure demandée dans la matrice pour créer un vecteur. La valeur la plus fortement représentée dans ce vecteur sera la première probabilité du positionnement à l'heure demandée par l'utilisateur. Par la suite, un deuxième vecteur plus petit, qui ne contient plus les données du premier vecteur est généré afin de prédire la deuxième probabilité. C'est exactement le même principe pour la troisième probabilité. Finalement, si plus de trois coordonnées différentes sont présentes, un pourcentage de valeur « autre » est affiché.

Voici les difficultés principales rencontrées :

- 1) Le programme permet de changer la variable de l'heure locale. Cela veut dire, qu'il faut reculer un certain nombre « X » de positions dans la matrice pour extraire les données. Cela en tenant compte de la précision du temps précédemment défini par l'utilisateur.

Explication avec un exemple :

L'heure locale à Genève pendant la période d'été est GMT+2 heures. Si l'utilisateur a défini une précision temporelle de la matrice d'une heure, alors c'est deux positions en arrière à partir de l'heure demandée par l'utilisateur que les données sont extraites. Par contre, si l'utilisateur a défini une précision de 15 minutes, alors il faut reculer de huit positions.

- 2) Synchronisation de l'heure demandée par l'utilisateur avec la précision de la matrice. En effet, l'utilisateur peut être amené à entrer n'importe quelle heure, pour une matrice avec une précision également variable.

Explication avec deux cas de figure :

- a) La matrice à une précision élevée alors que l'utilisateur demande la prédiction à l'heure près.
- b) L'utilisateur demande une heure très précise, comme par exemple à la seconde près, alors que la matrice elle, ne contient qu'une donnée par heure.

6.2 Implémentation en tenant compte des jours de la semaine

Le même principe que lors du sous-chapitre précédent est appliqué mais comme seule différence, la possibilité d'introduire le jour de la semaine.

Contrairement à l'algorithme qui ne tient pas compte des jours de la semaine (l'heure demandée par l'utilisateur est extraite à chaque ligne), le programme va prendre en compte seulement une ligne sur sept de la matrice pour déterminer la probabilité.

La difficulté principale, est de synchroniser le jour de la semaine demandé par l'utilisateur avec le jour de la semaine auquel l'enregistrement des données GPS a débuté. En voici un exemple (voir fig. 17) :

L'heure demandée est de 04:00 heures local (donc 02:00 heures GMT) pour les mardis :

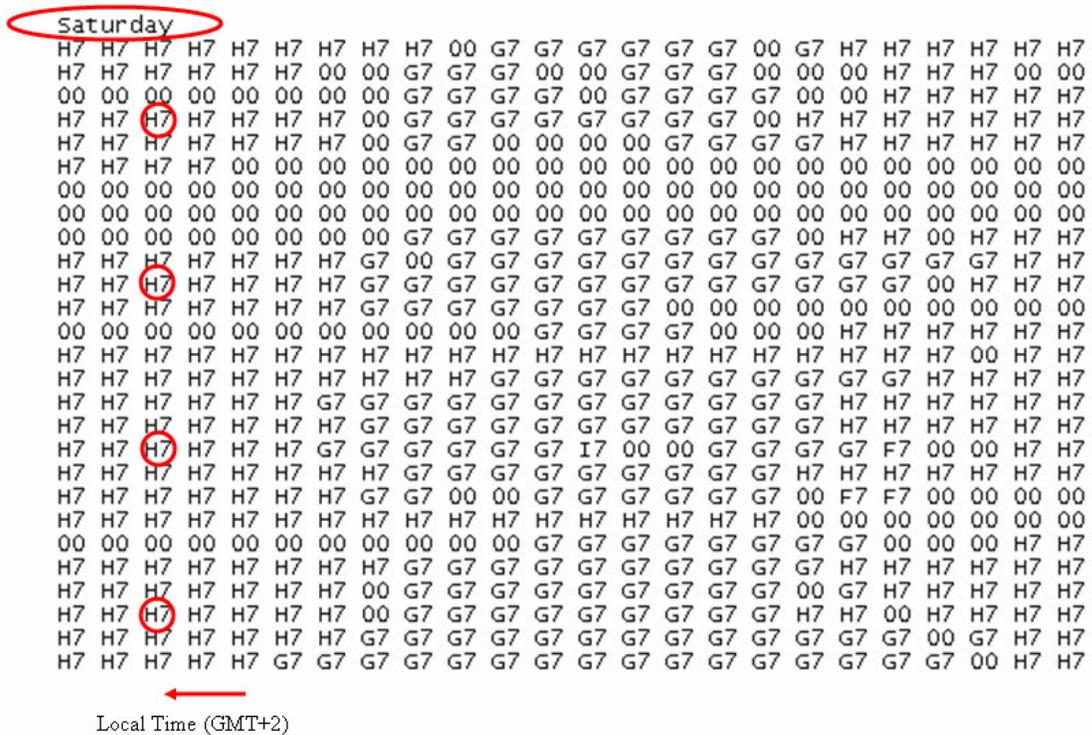


Fig. 17 : Exemple d'une matrice qui débute un samedi

Il faut donc que le programme aille chercher la bonne ligne au départ après l'interprétation de la première qui est dans ce cas le samedi. Cela, avant de procéder à l'incréméntation + 7.

Chapitre 7

Evaluation et cas d'utilisation du programme de prédiction

7.1 *Evaluation du programme de prédiction*

Le pourcentage de valeurs nulles

A ce stade du projet, j'ai accumulé 33 jours de données GPS (du jeudi 24 juillet 2008 au lundi 25 août 2008). Pour valider le programme de prédiction, toute la phase analytique du chapitre 5 est reprise dans ce chapitre, mais en utilisant mes données.

Un nouveau test concernant le pourcentage des valeurs nulles lors de l'enregistrement avec le Qtek est effectué. Le résultat, est de **24.12%** alors qu'il était de 33.88% lors de la première semaine. Le taux reste élevé mais il a quand même chuté de près de 10%. Ce résultat est très réjouissant. La fiabilité de la prédiction devrait sans aucun doute être plus élevée.

Composition du modèle d'une journée type

Pour vérifier le taux de corrélation, un modèle jour type a également été construit pour mes données.

Comme je travaillais du 24 juillet 2008 au 16 août 2008 généralement de 17:00 heures à 21:00 heures, une activité supplémentaire par rapport au test précédent a été ajoutée. Comme avant, on ne tient pas compte des déplacements. Uniquement les lieux fixes sont représentés lors de cette analyse, afin d'augmenter au maximum les probabilités de corrélations (voir fig. 18).

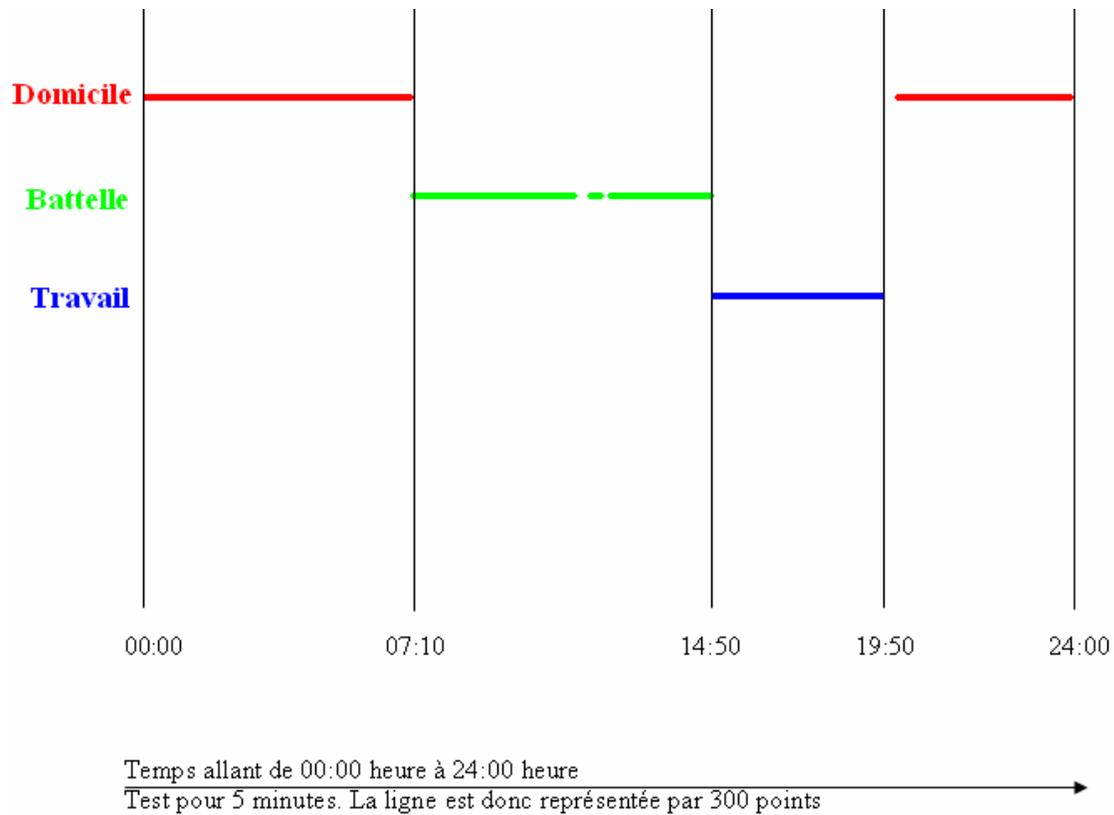


Fig. 18: Séquence d'heures dominantes

Conception du graphique :

Uniquement les valeurs du domicile, Battelle et du travail sont prises en compte, dans un souci de simplification et abstraction du model. Toutes les cinq minutes, une demande est effectuée concernant la valeur dominante.

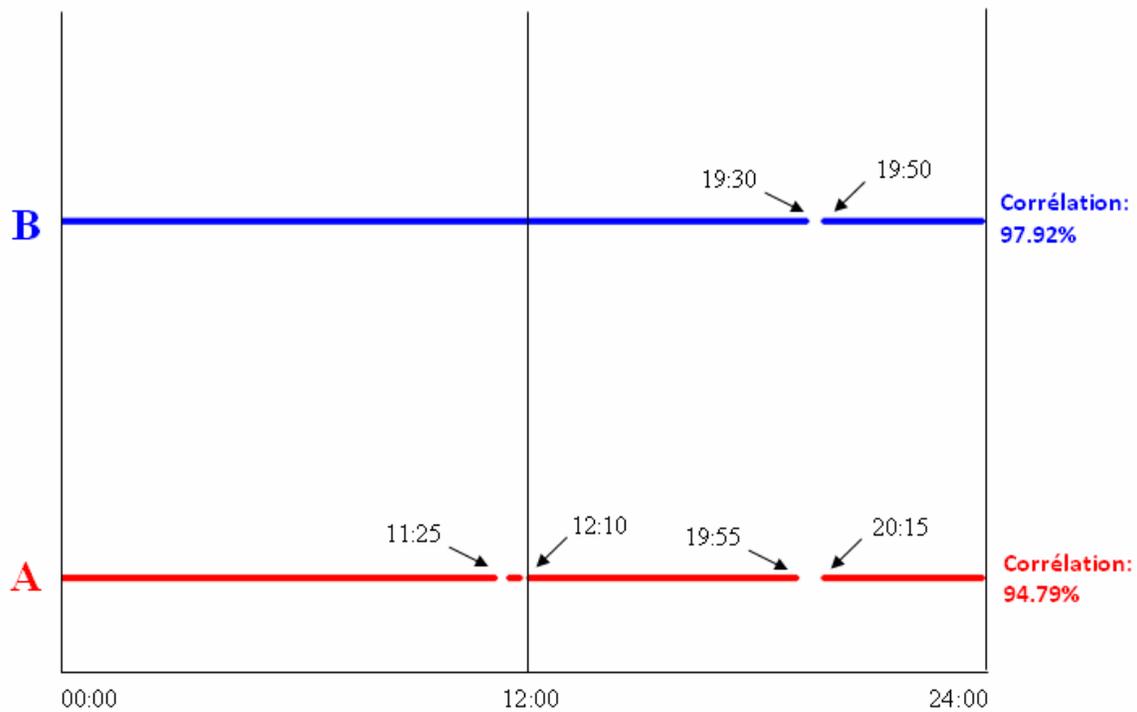
Lorsque la ligne n'est pas continue, comme vers 12:00 heures et 20:00 heures, cela signifie qu'aucune des trois données n'a été trouvées.

Attention, contrairement à l'analyse du chapitre cinq, ce test est effectué en GMT0.

Cette journée est donc composée de la façon suivante :

00 :00 – 08 :45	→	domicile
08 :45 – 16 :30	→	Battelle
16 :30 – 21 :30	→	travail
21 :30 – 24 :00	→	domicile

Là aussi, les deux algorithmes, celui remplaçant les valeurs nulles par les dernières données disponibles et celui qui tient compte des valeurs nulles, sont comparées (voir fig. 19).



Temps allant de 00:00 heure à 24:00 heure (de 0% - 100% de corrélation)
 Test pour 5 minutes. La ligne est donc représentée par 300 points

Fig. 19 : Comparaison de la corrélation avec le modèle de jour type des deux algorithmes

Analyse des périodes sans corrélation :

Attention, là aussi l'heure est exprimée en GMT0.

Les deux algorithmes présentent une corrélation parfaite lors du déplacement du domicile à Battelle. L'algorithme tenant compte des valeurs nulles présente une non corrélation lors de la pause de midi. Un second défaut, cette fois concernant les deux algorithmes, est constaté lors du retour au domicile.

Cela est notamment dû au fait que le Qtek se bloquait fréquemment lors des trajets. Remarquant ce problème quelques minutes après être arrivé à destination, explique le décalage de quelques minutes entre l'heure d'arrivée effective et le résultat présent.

Cette analyse confirme le bon fonctionnement des deux algorithmes. Comme lors du premier test, l'algorithme remplaçant les valeurs nulles par les dernières données disponibles, se montre plus performant. Même si cela est un peu plus nuancé lors de ce cas. Néanmoins, le choix d'avoir pris l'algorithme qui remplace les valeurs nulles se trouve être dans tous les cas le bon.

Corrélation avec le jour le plus improbable

Le test suivant, consiste à connaître le taux de corrélation le plus faible. Comme précédemment, un jour improbable aux corrélations est utilisé face au modèle type d'une journée (voir fig. 20). C'est le dimanche qui se prête le mieux pour de faibles corrélations, car je les ai tous passés à la maison lors des 33 jours d'enregistrements.

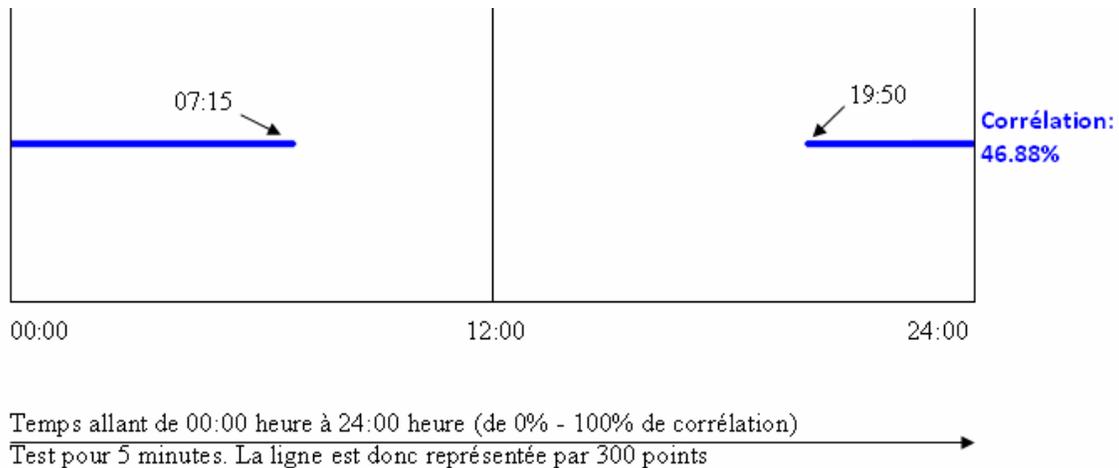


Fig. 20 : corrélation du modèle de jour type avec uniquement le dimanche

Encore une fois, le résultat est étonnement semblable à celui obtenu avec les données de Katarzyna Wac. Uniquement le temps passé le soir et le matin à la maison coïncide. Les données ne se recoupent pas durant la grande majorité de la journée. On détient la preuve, qu'en prenant en compte tous les jours de la semaine, la corrélation ne serait être satisfaisante pour des journées alternatives comme les week-ends par exemple. Le choix d'implémenter un algorithme qui tient compte des diverses journées de la semaine est donc totalement justifié.

7.2 Cas d'étude entre une résolution faible et une résolution élevée

Lors de ce sous-chapitre, une démonstration des divers cas d'utilisations est faite. Il s'agit de démontrer lors d'un premier exemple des situations facilement prédictibles. Dans un deuxième temps, on verra une condition où la prédiction est plus difficile à acquérir.

Cas d'étude avec un maillage et une précision temporels faible

Trois prédictions facilement à obtenir sont évaluées :
Ces testes ont été effectués avec une précision d'une heure et un maillage 10 x 10.

Il y a trois endroits qui sont des lieux où je me trouvais avec une très grande certitude lors des 33 jours d'enregistrement des données GPS pendant un certain laps de temps précis. Les trois prédictions concernent le lundi en tant que jour de la semaine.

Il s'agit des lieux suivant (voir fig. 21) :

- **I1 (X1)** se trouve à Bernex. C'est la commune où j'habite. La demande de prédiction est faite à 04 :00 heures du matin, dans le but d'être sûr que je sois à la maison. Sans surprise, le taux de probabilité est de 100% pour cette requête.
- **I6 (X2)** se trouve à Carouge (Battelle). C'est le lieu où je me rends pour faire mon travail de mémoire pour l'université. La requête de prédiction est faite pour 10:00 heures du matin. Comme auparavant, le résultat est de 100%. Là aussi, pas de surprise, en sachant que je m'y trouvais tous les matins hors week-end.
- **F6 (X3)** est le centre ville de Genève. Je m'y rends pour travailler en tant qu'employé à mi-temps au Quai du Sujet. Les tranches d'horaires que j'y effectue, sont variables. En principe, j'y travaille quatre jours sur cinq de 17:00 heures à 21:00 heures pendant la semaine. Par conséquent, la requête de la prédiction est faite pour 18:00 heures. Le résultat est le suivant : 75% pour X3 et 25% pour X2. L'origine du résultat est due au fait, que je ne travaillais pas le lundi 18 août 2008 et que je suis resté à Battelle.

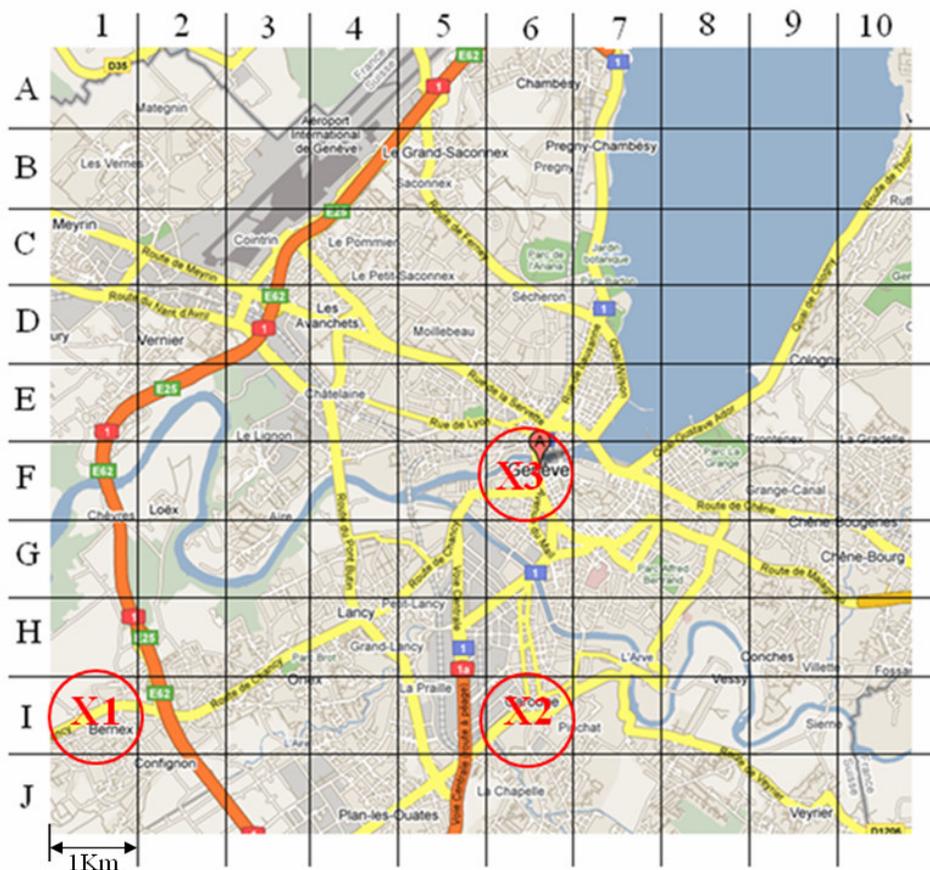


Fig. 21 : Carte de Genève (Google Maps)

Cas d'étude d'un déplacement avec un maillage et une précision temporels élevés

Ce cas de figure concerne les probabilités de positionnement lors d'une période de fortes chances de déplacement. C'est le trajet du lundi matin de mon domicile vers l'arrêt de bus de Plainpalais qui est pris comme cas d'étude. Je prends généralement le bus le matin entre 07:10 heures et 07:40 heures.

Ce test est effectué avec une précision de cinq minutes d'intervalle et un maillage 20 x 20.

Pour effectuer ce test, une prédiction toutes les cinq minutes est effectuée de 07:15 heures à 08:05 heures (voir fig. 22).

Le manque de données se fait ressentir lors de cet exemple. Il est évident, qu'en ayant seulement quatre jours à disposition (33 jours divisés par 7), les possibilités de bien évaluer une trajectoire sont minces. Cependant, comme on peut le constater sur la carte (fig. 23 de P1 – P8), la trajectoire du voyage est très visible.

7:15 AM	7:20 AM	7:25 AM	7:30 AM	7:35 AM	7:40 AM	7:45 AM	7:50 AM	7:55 AM	8:00 AM	8:05 AM
Q1 = 50.0% R1 = 50.0%	Q1 = 50.0% R1 = 50.0%	Q1 = 25.0% R2 = 25.0% R1 = 25.0% Autre 25.0%	P5 = 50.0% Q1 = 25.0% R2 = 25.0%	O8 = 50.0% Q1 = 25.0% Q4 = 25.0%	Q1 = 25.0% M10 = 25.0% P5 = 25.0% Autre 25.0%	K12 = 50.0% Q1 = 25.0% P5 = 25.0%	L12 = 50.0% Q1 = 25.0% P5 = 25.0%	Q1 = 25.0% M12 = 25.0% P5 = 25.0% Autre 25.0%	M12 = 50.0% Q1 = 25.0% O12 = 25.0%	M12 = 75.0% O12 = 25.0%
Pos.1 → R1	Pos.1 → R1	Pos.2 → R2	Pos.3 → P5	Pos.4 → O8	Pos.5 → M10	Pos.6 → K12	Pos.7 → L12	Pos.8 → M12	Pos.8 → M12	Pos.8 → M12

Fig. 22: tableau du positionnement lors du trajet Bernex – Plainpalais avec un intervalle de cinq minutes

Problèmes de données :

A 07:15 heures, il y a la même pondération entre Q1 et R1. C'est dû au fait que j'habite exactement à la frontière qui sépare ces deux positions (pour rappel, le GPS dispose d'une précision de 50 mètres en ville). Cette donnée est aléatoire et est susceptible de changer d'un jour à l'autre. Pour rendre ce cas d'étude plus compréhensible, on considère ces deux positions comme équivalentes. A 07:15 heures et 07:20 heures, je suis donc avec une probabilité de 100%, à la maison.

A 07:25 H / 07:40 H / 07:55 H, la pondération est la même entre les diverses données. J'ai pris arbitrairement en compte la valeur qui représente le mieux le déplacement.

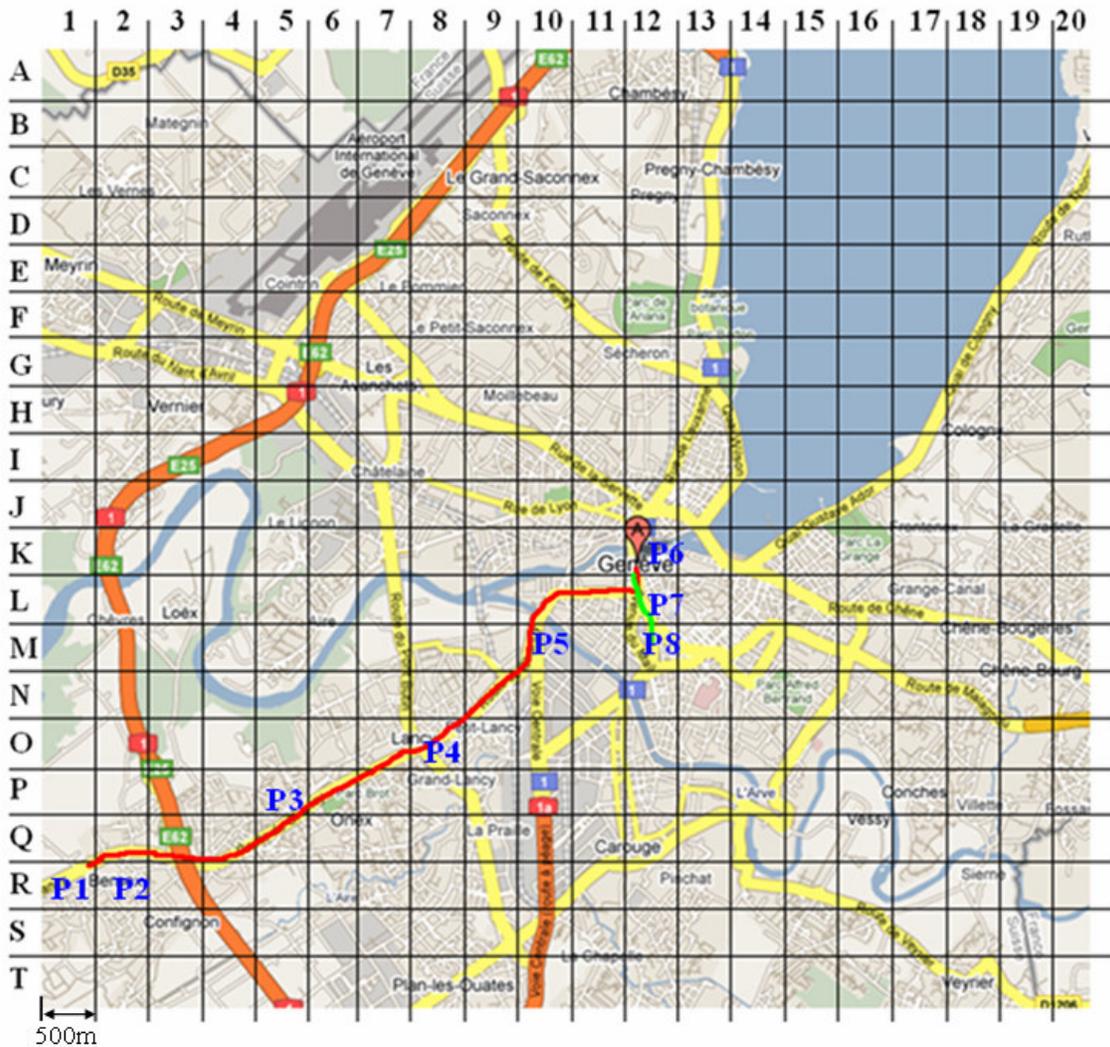


Fig. 23: Trajectoire de déplacement Bernex – Plainpalais (Google Maps)

Explication des lignes du graphique :

La ligne en rouge partant de Bernex direction centre-ville, s'effectue en bus. Le Trajet en vert est le déplacement à pied de Stand à Plainpalais.

Chapitre 8

Domaine d'utilisation

Comme mentionné dans certains passages de cette étude, les domaines d'utilisation de la prédiction géographique des personnes sont très divers et variés. Voici un résumé des applications possibles de ce projet :

Mobile Health Care

Le groupe de recherche Advanced Systems Group de l'Université de Genève consacre des efforts considérables pour apporter des solutions à l'assistance de personnes ayant des problèmes cardiovasculaires. Afin de permettre aux personnes souffrantes de ce type de syndromes de se déplacer plus librement dans leur environnement, des recherches sont effectuées pour signaler un éventuel danger au corps médical. Les patients sont équipés d'un électrocardiographe miniaturisé, permettant à tout moment de prendre connaissance d'un danger. Dans ce contexte, ce projet est d'une grande utilité dans trois situations bien précises.

Le premier champ d'application, consiste à relier l'information d'un problème cardiovasculaire ou d'épilepsie de façon fiable au corps médical (voir exemple de Sophie fig. 24). En cas normal, ce dernier est informé lors d'un problème via le réseau sans fil (comme par exemple 2.5G, 3G ou encore WLAN). Cependant, si la personne a un souci et ne se trouve pas dans un endroit couvert par ces réseaux, le programme de prédiction va intervenir. Grâce aux probabilités de positionnement que l'application génère, les ambulanciers auront des indices fiables pour retrouver la personne en danger.

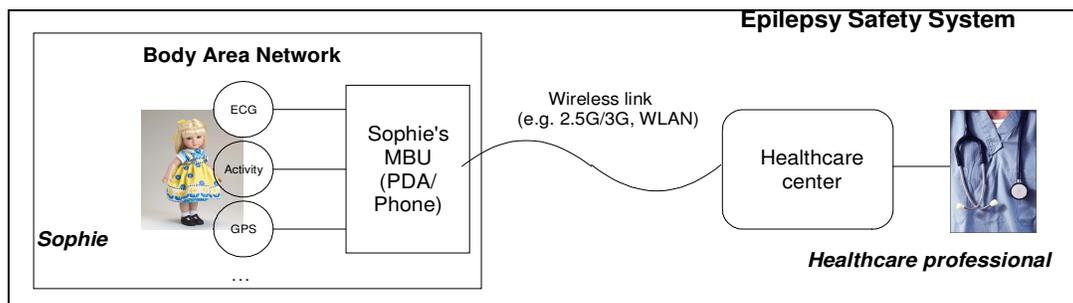


Fig. 24: Relais des données aux corps médical (source : thèse de Katarzyna Wac)

Un autre champ d'application, consiste à configurer le provider idéal sur l'appareil de transmission. Ces appareils sont équipés de smart card venant d'un pays autre que la Suisse. La connexion se fait donc en roaming, c'est-à-dire qu'à tout moment le provider peut être modifié. Advanced Systems Group possède de nombreuses statistiques

concernant ces providers (Swisscom et Sunrise). Les quality-of-services (puissance du réseau) sont connues pour toutes les localités et à tout instant. Grâce au système prédictif, l'appareil mobil pourrait être capable de choisir le provider le plus performant selon les statistiques présentes.

Finalement, le dernier champ d'application utilise la prédiction pour informer par sms des personnes volontaires à l'assistance aux personnes handicapées. C'est le système de prédiction qui va déterminer le choix de(s) personne(s) à contacter selon les probabilités de leur positionnement.

Télécom

Comme d'écrit lors du deuxième chapitre, les télécoms se montrent très intéressés par cette technologie. Le nombre d'applications multimédia nécessitant une bande passante conséquente se multiplie. Un système de prédiction pourrait être intégré dans chaque appareil mobil afin de réserver à l'avance de la ressource lors d'un changement de cellule.

Ce système serait d'un grand bénéfice pour le consommateur et le producteur. L'avantage de l'utilisateur serait d'avoir moins de coupures lors de lecture vidéo en streaming par exemple. Côté provider, l'avantage se situe dans la rationalisation des antennes, diminuant ainsi les investissements.

Cependant, le délicat problème des données confidentielles est posé. Si le système n'est pas bien protégé et que la surveillance des autorités ne se fait pas dans les règles, les opérateurs auront accès aux positionnements de leurs clients.

Dissémination/routage d'information géo-localisée sur des «mobile ad-hoc network»

Grâce à un service de prédiction de la localité futur des dispositifs mobiles, il serait possible de développer des algorithmes de routages et de dissémination d'information géo-localisés tel qu'il est envisagé dans le projet de recherche « *Hovering Information* » au sein du groupe ASG. Ces algorithmes profiteraient de ces prédictions pour choisir d'une manière plus intelligente les prochains dispositifs mobiles à transporter l'information d'un endroit à un autre.

Applications divers comme les réseau sociaux

En ayant la possibilité de prédire le lieu où vont se trouver ses connaissances, des applications très créatives deviennent imaginables.

Un exemple serait de confirmer un rendez-vous avec un collègue, si de très fortes probabilités subsistent de se trouver dans un futur proche au même endroit.

Une autre idée, serait de rendre l'appareil mobil plus autonome, en le laissant choisir le type d'alerte lors d'un appel. Cela, en fonction du lieu où se trouve l'utilisateur. Exemple : notification visuelle lors de cours à l'université, vibreur lors du travail au bureau ou sonnerie à la maison.

Chapitre 9

Conclusion

9.1 Généralités et objectif atteint

L'objectif de ce travail était de concevoir un programme capable de prédire le futur positionnement d'une personne grâce à un historique de données GPS. Ce but est atteint, comme cela a notamment été prouvé lors du septième chapitre.

Ce travail ne s'adresse cependant pas au grand public. La nécessité de collectionner des données GPS au préalable, en fait un programme qui n'est pas très commode à être utilisé. Cette étude s'adresse principalement au monde scientifique. Le fait que ce logiciel soit totalement paramétrable, lui donne la vertu d'être un outil dédié au domaine de la recherche.

9.2 Points faibles

Le problème principal est sans doute la lenteur d'exécution du programme. L'unique fait de parcourir l'ensemble du fichier des coordonnées GPS brutes (après un mois d'enregistrement, le fichier pèse plus de 155MB, donc ~5MB/jour), ralenti considérablement le logiciel. Le temps d'exécution du programme pour mes données avec un PC Pentium 4 cadencé à 3 GHz avec 1 GB de RAM tournant sur Windows XP, est de 12 secondes.

Même si ce travail s'adresse au monde de la science, une interface fait défaut. Défaut qui est principalement dû à un manque de temps. Comme les inputs et les outputs donnés par l'utilisateur, respectivement par le programme, sont peu nombreux, une interface serait réalisable dans un délai assez bref.

9.3 Futur travail

La prochaine étape sera, comme il était initialement prévu, de porter cette application sur mobil. Comme mentionné lors du sous-chapitre précédent, la rapidité d'exécution doit être nettement plus performante. Ce fait est particulièrement vrai pour un portage dédié aux appareils mobiles. Comme l'a suggéré l'assistant A. Villalba Castro, une solution serait de procéder avec une matrice qui se met à jour selon un intervalle de temps prédéfini, gardant ainsi seulement une fenêtre temporelle en mémoire.

Une autre évolution crédible du logiciel, consisterait en l'échange d'informations de prédiction entre usagers, afin qu'une interaction de plusieurs personnes soit possible, ceci dans l'esprit des applications sociales comme on a pu le voir lors du huitième chapitre.

Qui sait... peut-être qu'un jour, lors d'un travail de Master, j'aurai l'opportunité d'accomplir cette tâche avec à nouveau beaucoup de plaisir, tout comme cela a été le cas lors de la réalisation de ce projet !

Remerciements

Avant tout, j'aimerais remercier très sincèrement l'assistante Katarzyna Wac ☺☺☺. Sans elle, la réalisation de ce projet n'aurait pas pu aboutir à l'état actuel des choses. Par ailleurs, elle a toujours réussi à trouver le temps nécessaire lorsque j'avais besoin de son aide. J'apprécie ce fait d'autant plus qu'elle-même est sans doute très occupée à terminer sa thèse. J'aimerais profiter de cet instant, pour lui souhaiter bonne chance pour le travail qu'il lui reste à accomplir !

Mes remerciements vont bien entendu également à A. Villalba Castro. Il m'a beaucoup soutenu lors de la réalisation de ce projet. Par ailleurs, ses idées m'ont influencé très positivement pour la conception du travail.

A lui aussi, j'aimerais souhaiter bonne chance pour sa thèse !

Shaban Shaame m'a été d'une aide précieuse lors du début de ce projet. Il a pu m'apporter un soutien non négligeable pour me re-familiariser avec Java. A lui aussi j'aimerais souhaiter bonne chance pour la suite, notamment au sein de la start-up Organix (j'en profite pour également souhaiter bonne chance à Michel Deriaz à faire évoluer avec succès l'entreprise).

Mes sincères remerciements vont également à Michèle Frey, Alexandra Lauper, et Denise Kuhn. Merci, d'avoir eu la patience de relire à multiples reprises ce mémoire, afin de corriger les nombreuses fautes d'orthographe et autres tournures de phrases.

Bibliographie

- [USMP01] Chellappa-Doss, Jennings and Shenoy. *User Mobility Prediction in Hybrid and Ad Hoc Wireless Networks*
- [MMLC01] Tong Liu, Paramvir Bahl and Imrich Chlamtac. *Mobility Modeling, Location Tracking, and Trajectory Prediction in Wireless ATM Networks*
- [RPCD01] Kari Laasonen. *Route Prediction from Cellular Data*
- [MPMA01] Paolo Bellavista, Antonio Corradi and Carlo Giannelli. *Mobility Prediction for Mobile Agent-based Service Continuity in the Wireless Internet*
- [MDET01] Timothy Sohn, Alex Varshavsky, Anthony LaMarca and other. *Mobility Detection Using Everyday GSM Traces*
- [ULSL01] Daniel Ashbrook and Thad Starner. *Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users*
- [IDPT01] John Krumm and Eric Horvitz. *Predestination: Inferring Destinations from Partial Trajectories*
- [MPWN01] William Su, Sung-Ju Lee and Mario Gerla. *Mobility Prediction In Wireless Networks*
- [PWWT01] John Krumm and Eric Horvitz (Microsoft Research). *Predestination: Where Do You Want to Go Today*
Dérivé de la recherche suivante:
Predestination: Inferring Destinations from Partial Trajectories
Consultable sur le lien suivant:
<http://research.microsoft.com/users/jckrumm/predestination.pdf>

Webographie

Renseignements divers :

<http://fr.wikipedia.org/wiki/Wiki>

Trouver les coordonnées GPS :

<http://www.satsig.net/maps/lat-long-finder.htm>

Questions concernant la programmation en Java (*format jour de semaine*) :

<http://www.commentcamarche.net/forum/>

[MPEH01] Recherches sociologiques du comportement humain :

<http://news.bbc.co.uk/2/hi/science/nature/7433128.stm>

[ELBS01] Statistique du nombre de personnes abonnées au service GPS :

<http://www.cellular-news.com/story/19632.php>

<http://www.3g.co.uk/PR/Sept2006/3701.htm>

Diverses images :

http://www.switzernet.com/public/070608-subscribe/images/schema_BT100.png

http://assistance.orange.fr/Image/1509_schema_Install.png

<http://www.planetips.com/images/vsat.gif>

<http://www.123farmworks.com/emtactbgps.htm>

<http://www.unige.ch/ses/index.html>