

FACULTAT D'INFORMÀTICA DE BARCELONA (FIB),
UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC)—BARCELONATECH

Anàlisi d'algorismes en el traçat de rutes de senderisme a partir de dades GPS

GRAU EN ENGINYERIA INFORMÀTICA

Treball Final de Grau

Autor: DAVID DURAN ROSICH

Especialitat: *Computació*

Directora: VERA SACRISTÁN ADINOLFI

Matemàtiques Aplicades II

Codirector: RODRIGO I. SILVEIRA

Matemàtiques Aplicades II

1 de juliol de 2015

Resum

La *generació d'un mapa* consisteix a agrupar trajectòries GPS d'una zona per tal de recrear el seu mapa corresponent. Atès que aquest problema s'ha enfocat històricament en l'àmbit de les *ciutats*, la literatura en la generació de mapes de *senderisme* és escassa. En aquest projecte analitzem dos dels algorismes més recents (2012) pensats per *ciutat* aplicats en quatre zones diferents de *senderisme* de Catalunya; dues zones planes i dues de muntanya. A partir dels mapes generats hem analitzat el comportament dels dos algorismes i quins factors influeixen en la qualitat del seu resultat.

Resumen

La *generación de un mapa* consiste en agrupar trayectorias GPS de una zona con tal de recrear su mapa correspondiente. Dado que este problema se ha enfocado históricamente en el ámbito de las *ciudades*, la literatura sobre generación de mapas de *senderismo* es escasa. Este proyecto analiza dos de los algoritmos más recientes (2012) pensados para *ciudades*, y los aplica en cuatro zonas diferentes de *senderismo* en Cataluña; dos zonas planas y dos de montaña. A partir de los mapas generados se analiza el comportamiento de ambos algoritmos y qué factores influyen en la calidad de su resultado.

Abstract

Map construction refers to the process of grouping GPS traces from a certain zone with the purpose of building its corresponding map. Given that historically researchers have been interested only in map construction on cities and other urban areas, literature about map construction on hiking areas is scarce. This project analyses two of the most recent algorithms (2012) designed to work on cities, applied to four different hiking zones in Catalonia; two of them flat, and the other two mountainous. The obtained results are used to study the behaviour of both algorithms and to determine which factors influence their quality.

Agraïments

Vull agrair als meus directors Vera Sacristán i Rodrigo I. Silveira per la seva tutoria en aquest projecte.

Vull agrair a Enle Lin per les nostres converses fructíferes i la seva aportació de dades en el projecte.

Vull agrair a Mahmuda Ahmed, Sophia Karagiorgou, Dieter Pfoser i Carola Wenk per la seva aportació a la comunitat científica pel portal de referència en la generació de mapes mapconstruction.org.

Vull agrair a Nil Mamano i Andrés Mingorance per ajudar-me en la correcció de faltes.

Finalment, vull agrair el suport que m'han donat amics i familiars fins al dia d'avui.

Índex

1	Introducció	6
1.1	La nostra contribució	8
1.2	Estructura del treball	8
2	Estat de l'art i treballs relacionats	9
2.1	Algorismes per a la generació de mapes	9
2.1.1	Classificació	10
2.1.2	Ahmed i Wenk	11
2.1.3	Karagiorgou i Pfoser	15
2.2	Mesures de la qualitat	22
2.2.1	Distància Hausdorff dirigida	23
2.2.2	Distància basada en camins	23
2.2.3	Distància de camins més curts	24
3	Conjunts de dades	25
3.1	Preàmbul	25
3.2	Terrenys	25
3.2.1	Delta del Llobregat	27
3.2.2	Aiguamolls Baix Empordà	28
3.2.3	Turó de l'home	28
3.2.4	Garraf	29
3.3	Trajectòries	30
3.3.1	Tractament de les trajectòries	31
3.4	<i>Ground-truth</i>	33
3.4.1	Tractament dels <i>ground-truth</i>	34
4	Experimentació	39
4.1	Descripció de l'entorn de treball	39
4.1.1	Format de les dades	40

4.2	Parametrització dels algorismes	43
4.2.1	Identificació i extracció dels paràmetres	44
4.2.2	Cerca de valors per als paràmetres	46
4.2.3	Distinció angular de les interseccions	56
4.2.4	Mapes generats	61
4.3	Artefactes generats	65
4.3.1	Ahmed i Wenk	65
4.3.2	Karagiorgou i Pfooser	71
4.4	Punts forts i mancances dels algorismes en el senderisme	76
4.5	Mesures de qualitat	80
4.5.1	Distància Hausdorff dirigida i distància basada en camins	80
4.5.2	Distància de camins més curts	82
5	Gestió del projecte	85
5.1	Planificació temporal	85
5.1.1	Valoració d'alternatives i pla d'acció	92
5.2	Gestió econòmica (pressupost)	93
5.2.1	Identificació i estimació dels costos	93
5.2.2	Control de gestió	96
5.3	Sostenibilitat i compromís social	97
5.3.1	Matriu de sostenibilitat	97
5.3.2	Sostenibilitat econòmica	97
5.3.3	Sostenibilitat social	98
5.3.4	Sostenibilitat mediambiental	98
5.4	Relació amb l'especialitat de Computació	99
5.4.1	Assignatures relacionades	99
5.4.2	Adequació amb l'especialitat	99
5.4.3	Competències relacionades	99
5.5	Llicències i normatives	100
6	Conclusions	101
6.1	Conclusions dels resultats	101
6.2	Feina a realitzar en un futur	103
6.3	Conclusions acadèmiques i personals	104
	Referències	105

1 Introducció

La recent ubiqüitat de dispositius amb funcionalitat GPS (dispositius mòbils, dispositius GPS per cotxes, dispositius GPS en transport públic, ...) ha fet que emergís la necessitat de fer recerca sobre com aprofitar les dades que aquests ofereixen: les trajectòries GPS.

Una trajectòria GPS ve descrita per punts en l'espai, obtinguts mitjançant la tecnologia GPS, i mesures del moment en què cada punt ha sigut mostrejat, de manera que ordenant els temps es té una discretització de la trajectòria real (amb un cert error) que ha fet el dispositiu GPS (vegeu la [Figura 1.1](#)). El camp que estudia aquestes dades és el camp d'anàlisi de trajectòries, el qual tracta diversos problemes com la segmentació de trajectòries (p. e. classificar la dificultat dels trams d'una cursa), reconeixement de trajectòries (p. e. de firmes analògiques), o la *generació de mapes*, problema on s'emmarca aquest treball.

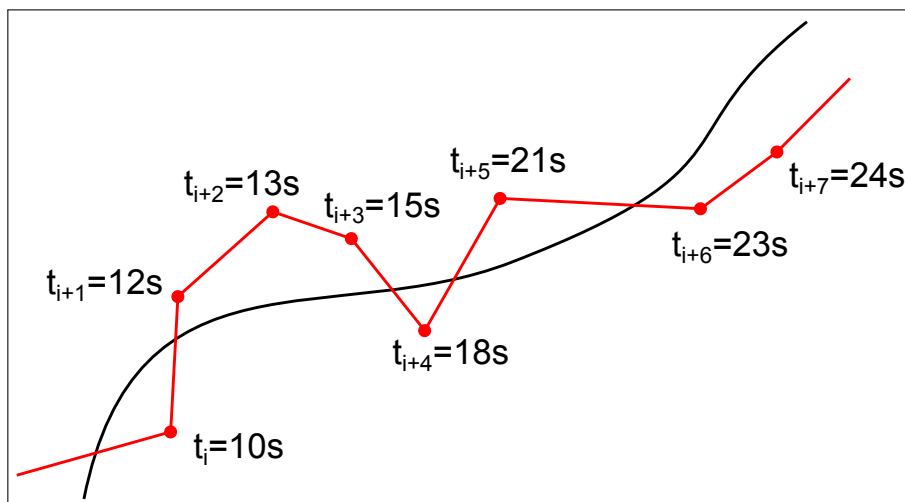


Figura 1.1: Els punts vermells són part d'una trajectòria GPS. Unint-los en el temps, se'ns mostra una aproximació de la trajectòria real (en negre).

El problema de la generació de mapes a partir de trajectòries GPS és un problema de notòria importància, ja que presenta nombroses utilitats com la construcció de mapes en zones poc cartografiades, la identificació d'incidències i camins tallats, l'actualització de mapes davant l'aparició de nous camins, etcètera. La generació de mapes parteix del fet que les trajectòries GPS aporten informació dels camins subjacents per on passen, però amb un cert error donat per la inexactitud de la tecnologia GPS. Per tant, el problema de la generació de mapes consisteix en, donat un conjunt de trajectòries que contenen error, reconstruir els camins per on aquestes passen així com les interseccions entre aquests camins, és a dir, acabar generant un mapa que cobreixi tota la zona mostrejada pel conjunt de trajectòries.

En els darrers anys, la comunitat científica s'ha centrat en la *generació de mapes de carrers i carreteres de ciutat* [2, 5, 12, 14, 15, 18, 19, 24, 28, 30, 31, 33, 35, 38, 39, 40] a partir de trajectòries GPS obtingudes per vehicles (transport públic, vehicles privats, ...). Un repte de cabdal importància és ampliar l'àmbit de la generació de mapes cap a altres tipus de mapes fora de ciutat, com són els *mapes en el senderisme*. El motiu d'aquest repte és la gran diferència que existeix en les trajectòries GPS d'un vehicle en una ciutat respecte les trajectòries d'un senderista, així com les diferències topogràfiques que presenten els mapes d'una ciutat respecte els mapes de zones pròpies del senderisme.

En el nostre saber, hi ha hagut pocs intents per part de la comunitat científica per resoldre la generació de mapes en l'àmbit del senderisme, i cap d'ells no ha aconseguit resultats suficientment satisfactoris. A més, no s'ha fet cap anàlisi de les diferències entre el problema de la generació de mapes en ciutat amb el de mapes en senderisme, ni des d'un punt de vista de l'adequació dels algorismes existents, ni tampoc de les diferències en les trajectòries i les consideracions que aquestes diferències comporten.

El treball aquí present, consisteix a efectuar una parametrització al màxim d'adequada per diferents dades de senderisme de dos algorismes ben coneguts en la generació de mapes per ciutat (Ahmed i Wenk [4] i Karagiorgou i Pfoer [31]), i analitzar els factors que ocasionen un detriment en la qualitat dels mapes generats. Considerem que aquesta anàlisi és indispensable per dues raons:

- Primerament, permet identificar les diferències que presenten els problemes de la generació de mapes en ciutat respecte la generació de mapes en senderisme des d'un punt de vista empíric.

- En segon lloc, permet marcar un punt de partida per a la recerca en la generació de mapes en el senderisme, atès que crea una classificació d'idees prèvies ja provades en la comunitat científica segons el seu grau d'adequació al senderisme.

1.1 La nostra contribució

En aquest treball analitzem, l'aplicació a trajectòries de quatre zones de senderisme de Catalunya, de dos dels algorismes més representatius i recents de l'estat de l'art en la generació de mapes per ciutat: *Ahmed i Wenk [4]* (2012) i *Karagiorgou i Pfoser [31]* (2012).

Les quatre zones de senderisme són el *Turó de l'home* (parc natural del Montseny), *Aiguamolls del Baix Empordà* (a la desembocadura del riu Ter), el *Garraf* i el *Delta del Llobregat*. Conjuntament amb les trajectòries també s'inclouen mapes reals digitalitzats de les zones respectives, obtinguts de OpenStreetMap i corregits per cobrir més terreny (referits d'ara en endavant amb el terme anglès *ground-truth*).

L'anàlisi que proveïm consisteix en un estudi aprofundit dels nostres conjunts de dades comparant-los amb les trajectòries de Ahmed et al. [5], la parametrització justificada dels dos algorismes i la identificació i classificació dels punts forts i mancances d'ambdós algorismes quan treballen en dades de senderisme.

Addicionalment a l'anàlisi i justificacions per inspecció visual, seguint el model de Ahmed et al. [5], comparem el resultat dels algorismes amb els *ground-truth* mitjançant tres mesures: distància dirigida Hausdorff [7], distància basada en camins [3] i distància de camins més curts [31].

1.2 Estructura del treball

En el **Capítol 2** s'expliquen els diferents algorismes existents per a la generació de mapes, explicant en detall els dos algorismes que analitzem, així com els mètodes de comparació entre un mapa generat i el seu *ground-truth*. Els conjunts de dades que emprem en els experiments s'expliquen exhaustivament al **Capítol 3**. L'anàlisi completa dels algorismes aplicats a trajectòries de senderisme es troba al **Capítol 4**. Previ al final, en el **Capítol 5**, s'explica com s'ha gestionat aquest treball i, per acabar, al **Capítol 6** hi ha les conclusions i pautes per a treballs futurs.

2 Estat de l'art i treballs relacionats

Al llarg de la literatura en la *generació de mapes*, no hi ha hagut casos, en el nostre saber, d'intentar comparar la qualitat dels algorismes actuals pensats per a *carrers i carreteres de ciutat* a dades d'altres zones com el ciclisme, l'escalada o el *senderisme*. Encara més, la comparació dels diferents algorismes a un conjunt de dades unificat no ha sigut una realitat fins la publicació del treball de Biagioni i Eriksson [11] (2012) on s'ha fet una comparació quantitativa entre 11 algorismes de generació de mapes [2, 4, 14, 15, 18, 19, 28, 30, 35, 38, 39, 40], posant a disposició pública tant les implementacions d'aquests com el seu conjunt de dades (*Chicago*). Els seus esforços han sigut recentment ampliat per Ahmed et al. [5] (2015), que han comparat el funcionament de 7 algorismes [12, 14, 18, 19, 24, 31] sobre quatre conjunts de dades (*Chicago*, *Atenes gran*, *Atenes petit* i *Berlín*).

Aquest capítol està dividit en dues seccions. Primerament, a la **Secció 2.1** s'expliquen els diferents enfocaments dels algorismes de la *generació de mapes* i, en més detall, els dos algorismes que s'analitzen: Ahmed i Wenk [4] i Karagiorgou i Pfoser [31]. Finalment, a la **Secció 2.2**, s'expliquen les tres mesures de qualitat per a mapes generats que hem utilitzat en l'experimentació.

2.1 Algorismes per a la generació de mapes

Els algorismes per a la *generació de mapes* consideren com a entrada un conjunt de *trajectòries*. Aquestes trajectòries vénen descrites per un seguit de *mostres* que consisteixen en un punt en el pla (*i. e.* amb coordenades (x, y)) i una marca de temps del moment en què la mostra ha estat presa. Generalment, les trajectòries es modelen com a línies poligonals que uneixen els punts de les mostres seqüencialment en el temps.

La sortida d'aquests algorismes és un *mapa*. Malgrat es pugui modelar un mapa de moltes formes, una de les més senzilles en la qual ens basarem és modelar-lo com un graf geomètric no dirigit, on cada vèrtex és un punt del pla i cada aresta una línia poligonal

que connecta dos vèrtexs. Una altra forma d'expressar aquest mateix model és considerar que les arestes són segments que connecten dos vèrtexs i que hi ha tres tipus de vèrtexs, els de grau 2 (vèrtex intermedis), els de grau més gran que 2 (interseccions) i els de grau 1 (vèrtexs extrems). A la [Figura 2.1](#) es mostra un exemple d'aquest model i com aquest permet aproximar amb una precisió arbitrària camins sinuosos simplement augmentant el nombre de vèrtexs intermedis.

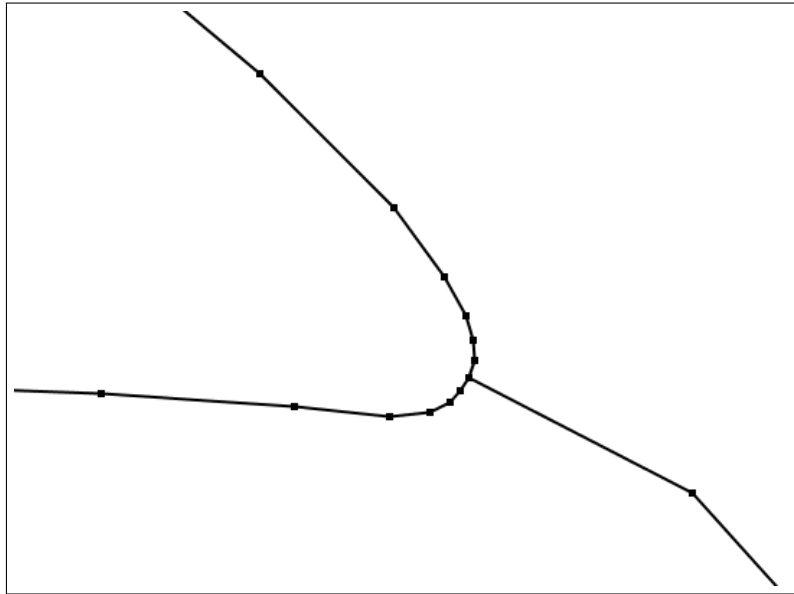


Figura 2.1: Representació visual del model d'un *mapa* en forma de graf geomètric.

2.1.1 Classificació

Seguint la classificació descrita en Ahmed et al. [5], els algorismes de generació de mapes es poden dividir en tres grups segons el seu enfocament del problema: *creació de clústers de punts*, *inserció incremental de trajectòries* i *unió d'interseccions*.

Creació de clústers de punts

Els algorismes d'aquesta categoria agafen de les trajectòries d'entrada un conjunt de punts, ja siguin els punts originals de les trajectòries o extraient-ne de nous, i en fan diversos clústers que finalment connecten entre ells. Per fer la construcció dels clústers de punts hi ha hagut tres tendències diferents:

- Utilitzar l'algorisme de k -means basat en mesures de distància com podria ser la Euclidiana. Alguns exemples d'aquest tipus d'algorisme són [19, 28, 40].
- A partir de mètodes KDE (de l'anglès, *Kernel Density Estimation*), obtenen mesures de la densitat dels punts les quals serveixen, per exemple, per eliminar *outliers* o per tenir punts amb diferents pesos segons la seva importància. Alguns exemples d'aquest tipus d'algorisme són [12, 15, 18, 33, 39].
- Utilitzar propietats de veïnatge local descrits per models propis de la geometria computacional, com els diagrames de Voroni [16, 24] o altres veïnatges més complexos com el complex Vietoris-Rips [1].

Inserció incremental de trajectòries

Els algorismes d'aquesta categoria generen el mapa a base d'inserir incrementalment les trajectòries a un mapa inicialment buit. La majoria d'aquests algorismes utilitzen idees de *map-matching* a l'hora de decidir com inserir les noves trajectòries. A vegades, altres informacions derivades com la distància i la direcció de la trajectòria es tenen en compte a l'hora d'inserir o eliminar trossos de trajectòries o del mapa construït fins el moment. En aquesta categoria hi entra l'algorisme Ahmed i Wenk [4], així com [13, 14, 35].

Unió d'interseccions

Encara que està relacionat amb la creació de clústers de punts, la idea darrera la unió d'interseccions és determinar primer quines són les interseccions en el mapa i, en un segon pas, unir-les. En aquesta categoria hi cau l'algorisme Karagiorgou i Pfoser [31], així com [21].

2.1.2 Ahmed i Wenk

L'algorisme de Ahmed i Wenk [4] (2012) actua per inserció incremental de trajectòries, és a dir, parteix d'un graf inicial buit i afegeix trajectòries fins a tenir el graf final que serà el mapa generat. La idea que hi ha al darrera és la d'agrupar les trajectòries per un criteri de proximitat espacial, el qual ve definit per un paràmetre de tolerància anomenat ϵ d'ara en endavant. Aquesta «proximitat» o «similitud» la mesuren mitjançant la

distància Fréchet δ_f que es defineix de la manera següent. Donades dues corbes al pla $f, g: [0, 1] \rightarrow \mathbb{R}^2$,

$$\delta_F(f, g) = \inf_{\alpha, \beta} \max_{t \in [0, 1]} \{d(f(\alpha(t)), g(\beta(t)))\} ,$$

on α, β són reparametrizacions de t (*i. e.* $\alpha, \beta: [0, 1] \rightarrow [0, 1]$) monòtones creixents i d denota distància entre dos punts, en aquest cas, distància Euclidiana. Informalment, si hom pensa de t com a temps, $f(\alpha(t))$ com la posició d'un gos i $g(\beta(t))$ la posició del seu amo, la distància Fréchet $\delta_F(f, g)$ és la mínima longitud (en termes de d) que ha de tenir la corretja perquè l'amo recorri tot el camí g i el gos tot el camí f sense que cap dels dos faci mai enrere.

Un concepte important per al càlcul de la distància Fréchet és el conjunt, anomenat en anglès *free space* (espai lliure):

$$F_\varepsilon(f, g) := \{(s, t) \in [0, 1]^2 \mid d(f(s), g(t)) \leq \varepsilon\}, \varepsilon > 0.$$

Una de les propietats descrites per Alt i Godau [6] és que $\delta_F(f, g) \leq \varepsilon$ si, i només si, existeix una corba en F_ε monòtona en ambdues coordenades que va des de la cantonada inferior esquerra (*i. e.* el punt $(0, 0)$) a la cantonada superior dreta (*i. e.* el punt $(1, 1)$). Partint el quadrat $[0, 1]^2$ en les zones pertanyents a F_ε i les que no, s'obté el diagrama FD_ε el qual permet il·lustrar el concepte de F_ε així com la propietat anterior (vegeu la [Figura 2.2](#)). Finalment, Ahmed i Wenk [4] defineixen el concepte de $FS_\varepsilon(G, l)$, per a un graf $G = (V, E)$ i una corba l , com una col·lecció de diagrames $FD_\varepsilon(e, l)$ per totes les arestes $e \in E$ agrupades segons la informació d'adjacència de G .

Algorisme

L'algorisme de Ahmed i Wenk [4] es divideix en dues fases. La primera fase consisteix en fer un *map matching* d'una corba l_i en el mapa incremental G_{i-1} amb l'ajuda de FS_ε . Seguidament, la segona part consisteix en actualitzar el mapa incremental G_{i-1} . Per una banda, s'hi afegixen les subcorbes de l_i que no estan suficientment aprop de G_{i-1} . Per una altra, es modifiquen les arestes e properes a alguna subcorba de l_i amb la finalitat de simplificar la geometria de e i, per tant, de G . L'[Algorisme 2.1](#) mostra aquest algorisme a alt nivell.

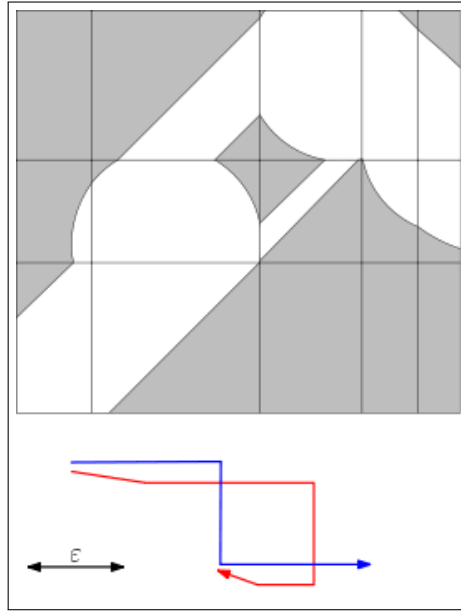


Figura 2.2: $FD_\varepsilon(f, g)$ de les corbes vermella i blava. En blanc es mostra F_ε . Com es pot observar, aquestes dues corbes tenen distància Fréchet més gran que ε ja que la zona propera a $(1, 1)$ està fora de F_ε , és a dir, les corbes divergeixen més que ε cap el seu final. La idea al darrere del diagrama FD_ε és que qualsevol corba l de FD_ε monòtona en ambdues direccions que vagi des de $(0, 0)$ a $(1, 1)$ i que passi només per la zona blanca ens permet definir unes parametritzacions $\alpha(t) = l_x(t)$ i $\beta(t) = l_y(t)$ tals que $\max_{t \in [0, 1]} \{d(f(\alpha(t)), g(\beta(t)))\} \leq \varepsilon$ i, per tant, l'ímfim d'aquesta funció sobre totes les α, β , *i. e.* la distància Fréchet entre f i g , també serà menor a ε .

Algorisme 2.1 Ahmed i Wenk [4].

```

1:  $G_0 \leftarrow \text{EMPTY}$ 
2: for all  $l_i \in \text{INPUT}$  do
3:   Calcula map-matching  $M_{1.5\varepsilon}(l_i, G_{i-1})$ 
4:   for all  $(l_i[a_{j-1}, a_j], X) \in M_{1.5\varepsilon}(l_i, G_{i-1})$  do
5:     if  $X$  és NULL then                                 $\triangleright l_i[a_{j-1}, a_j]$  no està associat a cap  $e \in E$ 
6:        $G_i :=$  inserir  $l_i[a_{j-1}, a_j]$  a  $G_{i-1}$ 
7:     else                                                 $\triangleright l_i[a_{j-1}, a_j]$  està associat a  $e[\text{ini}, \text{fi}]$  on  $e \in E$ 
8:        $G_i :=$  modifica  $e[\text{ini}, \text{fi}]$  de  $G_{i-1}$  a partir de  $l_i[a_{j-1}, a_j]$ 
9:     end if
10:  end for
11: end for

```

Els detalls d'aquest algorisme es poden trobar a Ahmed i Wenk [4]. Tanmateix, a continuació se'n fa una explicació resumida de cada un dels passos:

Map-matching $M_\gamma(l, G)$ **d'una corba l en un graf G** La idea d'un *map-matching* $M_\gamma(l, G)$ és trobar subcorbes de l que estiguin en la zona blanca de $FS_\gamma(G, l)$, *i. e.* tals que la distància Fréchet de la subcorba a una porció $[ini, fi]$ d'una aresta e de G , és menor a γ . El resultat d'aquest *map-matching* són les diferents parelles $(l_i[a_{j-1}, a_j], X)$ on $0 \leq a_0 < a_1 < \dots < a_k \leq 1$ i X és $e[ini, fi]$ si $l_i[a_{j-1}, a_j]$ es troba a distància Fréchet menor o igual a γ de $e[ini, fi]$ o, en cas que no hi hagi cap aresta que ho compleixi, $X = \text{NULL}$. S'assumeix que les subcorbes $l_i[a_{j-1}, a_j]$ són maximals tant si tenen aresta associada com si estan a la zona grisa de $FS_\gamma(G, l)$. Cal observar que en l'algorisme original aquest *map-matching* es fa amb valor $\gamma = 1.5\varepsilon$.

Inserció d'una corba en un graf G Es parteix d'una subcorba $l_i[a_{j-1}, a_j]$ que s'insereix directament a G . Finalment, per construcció, les subcorbes anterior i posterior de $l_i[a_{j-1}, a_j]$ són o bé zones associades a una aresta de G o bé a_{j-1} i/o a_j són extrems de la corba (i en tal cas no cal fer res). Suposant que a_{j-1} no sigui un extrem (es fa simètricament per a a_j si aquest no és un extrem) i que la subcorba anterior $l_i[a_{j-2}, a_{j-1}]$ està associada a $e_{prev}[ini, fi]$, llavors es connecta la subcorba $l_i[a_{j-1}, j]$ a $e_{prev}[fi]$ de manera que s'afegeix el segment $e_{prev}[fi]l_i[a_{j-1}]$ com a aresta de G (en cas que $e_{prev}[fi]$ no fos un vèrtex original de G , aquest se li afegeix).

Modificació d'una aresta $e[ini, fi]$ a partir d'una corba $l_i[a_{j-1}, a_j]$ La idea d'aquesta modificació és reduir la complexitat de G sense comprometre els llindars de qualitat del mapa generat. Per fer-ho, s'agafa una seqüència de vèrtexs γ' de $e[ini, fi]$ i de $l_i[a_{j-1}, a_j]$ seguint una corba poligonal monòtona en $FD_{1.5\varepsilon}(e[ini, fi], l_i[a_{j-1}, a_j])$. Arribats a aquest punt, s'aplica l'algorisme d'enllaç mínim [27] a les boles centrades als diferents vèrtexs de γ' i radi 2ε i s'obté un camí amb la mateixa fita en la distància Fréchet respecte el camí real però de complexitat igual o menor al camí real, sempre i quan es compleixin les hipòtesis presentades a continuació.

Hipòtesis

En l'article original [4] s'inclouen un seguit d'hipòtesis sobre les dades i el mapa real de la zona les quals simplifiquen el càlcul de l'algorisme i garanteixen una qualitat teòrica del mapa generat. Les hipòtesis són quatre:

1. Cada camí del mapa real té alguna porció separada per una distància al menys 3ϵ de qualsevol altre camí.
2. Si dos camins del mapa real s'apropen a una distància $\leq 3\epsilon$, aleshores comparteixen un vèrtex v .
3. L'error de les trajectòries respecte el camí que mostregen és més petit que $\epsilon/2$.
4. Cada trajectòria fa un recorregut acíclic en el mapa real de la zona.

En el cas que les hipòtesis siguin certes, es garanteix que el mapa generat preserva tots els camins del mapa real mostrejats per les trajectòries. Aquí *preservar* un camí significa que totes les seccions del camí en el mapa real que estiguin més lluny de 3ϵ de qualsevol altre camí («bones seccions»), seran representades per un camí del mapa generat que tindrà com a molt distància Fréchet ϵ respecte al real en aquestes «bones seccions».

2.1.3 Karagiorgou i Pfoser

L'algorisme de Karagiorgou i Pfoser [31] (2012) funciona per detecció i connexió d'interseccions, és a dir, identifica les interseccions del mapa i llavors les uneix mitjançant la informació de les trajectòries. Aquest algorisme està dividit en quatre fases: *identificació d'interseccions*, *enllaç d'interseccions*, *correcció dels enllaços* i un post-processat d'*eliminació d'interseccions triangulars*. La descripció que es dona d'aquest algorisme és una adaptació de Karagiorgou i Pfoser [31], on hi podeu trobar més detalls.

Identificació d'interseccions

La detecció d'interseccions es fa per agrupació de girs que prèviament han sigut identificats. Un cop identificats aquests girs, aquests s'agrupen segons proximitat i tipus de gir i, finalment, aquestes agrupacions de girs es tornen a agrupar per formar les

interseccions. El pseudocodi d'aquesta fase el podeu trobar a [Algorisme 2.2](#). Cal dir que aquesta fase utilitza quatre paràmetres que anomenarem *angle gir*, *màx. velocitat gir*, *dist. girs* i *dist. inter.*, a fi de poder-los identificar en seccions posteriors.

La *identificació dels girs* es realitza mitjançant dos indicadors derivats del sentit comú del moviment de vehicles: primerament, els vehicles redueixen la seva velocitat i, en segon lloc, canvien la seva direcció quan efectuen un gir. Considerant dos llindars, un per la desviació angular (*angle gir*) i un altre per la velocitat (*màx. velocitat gir*), totes les posicions dins de cada trajectòria que els satisfan són identificades com a girs.

Un cop identificats els girs, aquests s'*agrupen* a partir de la seva similitud angular i proximitat espacial. La similitud angular de dos girs ve donada per un model de girs en els quals s'identifiquen vuit tipus de girs segons la seva orientació absoluta respecte els punts cardinals, de manera que els girs provinents del sud dirigits a l'est es marquen com a 1, girs d'est a sud com a 2, girs d'oest a sud com a 3, etcètera (vegeu la [Figura 2.3](#)). El criteri de proximitat espacial ve definit pel llindar *dist. girs*.

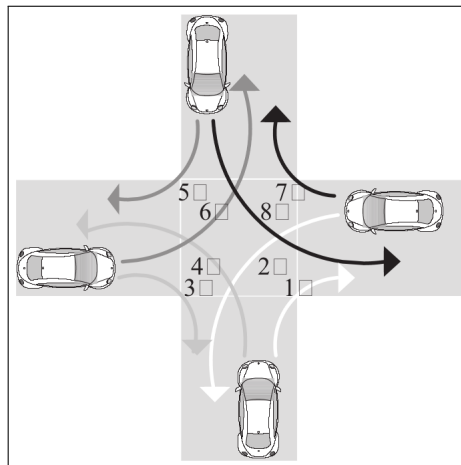


Figura 2.3: Model de girs de Karagiorgou i Pfoser [31]. Imatge extreta de l'article original.

Finalment, a partir de les agrupacions de girs, aquestes es tornen a agrupar per un criteri espacial i acaben esdevenint *interseccions*. A l'hora de fer aquesta segona agrupació s'enregistren dues informacions, la primera és el nombre de girs originals amb els quals s'ha derivat la intersecció i, la segona és quins tipus de girs (seguint el model de girs explicat) «permet» la intersecció, és a dir, la unió de tots els tipus de girs dels quals la intersecció conté com a mínim un exemple. El criteri espacial d'aquesta segona agrupació ve definit a partir del llindar *dist. inter.*

Algorisme 2.2 Identificació d'interseccions.

```

1: function INTERSECCIONS( $T$ , angle gir, màx. velocitat gir, dist. girs, dist. inter.)
2:    $P_S \leftarrow \emptyset$ 
3:   for all  $t \in T$  do ▷ per totes les trajectòries
4:     for all  $(p_{prev}, p, p_{seg})$  ternes de punts consecutius en  $t$  do
5:        $\alpha_p \leftarrow \text{DIFANGULAR}(p_{prev}, p, p_{seg})$ 
6:        $v_p \leftarrow \text{VELOCITAT}(p_{prev}, p)$  ▷ Velocitat d'anar des de  $p_{prev}$  a  $p$ 
7:       if  $(\alpha_p \geq \textit{angle gir} \wedge v_p \leq \textit{màx. velocitat gir})$  then
8:          $P_S.\text{INSERT}(p, \text{TIPIUSGIR}(p))$ 
9:       end if
10:    end for
11:  end for
12:   $P_C \leftarrow \text{AGRUPAPUNTS}(P_S, \textit{dist. girs})$  ▷ Genera agrupacions de girs
13:   $I \leftarrow \text{AGRUPAPUNTS}(P_C, \textit{dist. inter.})$  ▷ Genera interseccions
14: end function

```

Enllaç d'interseccions

En aquest punt de l'algorisme, es tenen calculades les interseccions de forma aïllada. Aquesta fase consisteix en trobar formes d'unir les interseccions mitjançant *enllaços* derivats de les trajectòries. El pseudocodi per aquesta fase la podeu trobar a l'[Algorisme 2.3](#).

Primerament, es consideren totes les trajectòries que han originat més d'una intersecció a partir dels seus girs. Llavors, per cada una d'aquestes trajectòries s'enllacen les dues (o més) interseccions que origina i s'associa a cada l'enllaç els trossos de la trajectòria que uneixen les interseccions.

Un cop finalitzat el pas anterior, es té un graf d'interseccions connectades per enllaços, on cada enllaç té un o diversos trossos de diferents trajectòries. En aquest punt convé agrupar els diferents trossos de trajectòries per acabar generant un enllaç incrustat (en forma de línia poligonal) entre les diferents interseccions. Per agrupar els trossos de trajectòria es fa servir un algorisme d'escombrat (en anglès *sweep-line algorithm*) en el conjunt de punts de tots els trossos de trajectòries on, per cada un dels punts, es calcula una posició mitjana basada en la distància normal del punt a la resta de trajectòries. Finalment, l'enllaç queda representat per la unió de tots aquests punts mitjana. A la [Figura 2.4](#) es mostra un esquema de com funciona l'algorisme d'escombrat.

Arribats a aquest moment, tots els enllaços entre trajectòries estan representats per una única línia poligonal. A més, durant la construcció de la línia poligonal s'associen a

Algorisme 2.3 Enllaç d'interseccions.

```

1: function ENLLAÇ( $T, I$ )
2:    $IS \leftarrow \emptyset$ 
3:   for all  $t \in T$  do                                     ▷ per totes les trajectòries
4:     for all  $p \in t$  do                                     ▷ per tots els punts de la trajectòria
5:       if  $p \in I$  then
6:          $IS \leftarrow IS \cup \{i^-, i, p^-, p, t\}$  ▷ es guarden les dues interseccions  $i^-$  i  $i$  i
           les seqüències de punts entre ambdues
7:       end if
8:     end for
9:   end for
10:   $L \leftarrow \emptyset$ 
11:  for all  $l = \{i^-, i, \dots\} \in IS$  do                       ▷ per tots els enllaços
12:     $LS \leftarrow \emptyset$ 
13:    for all  $t \mid \{i^-, i, p^-, p, t\} \in IS$  do             ▷ per totes les trajectòries associades
14:       $LS \leftarrow LS \cup \{t, p^-, \dots, p\}$  ▷ es guarda el tros de trajectòria corresponent
15:    end for
16:     $l.numtraj \leftarrow \text{MIDA}(LS)$ 
17:     $(l, l.distmax) \leftarrow \text{ESCOMBRAT}(LS)$  ▷ agrupa trossos de trajectòries i guarda
           distància màxima
18:     $L \leftarrow L \cup \{l\}$ 
19:  end for
20: end function

```

Correcció dels enllaços

En la fase anterior s'han tractat tots els trossos de les trajectòries que unien les interseccions. Tanmateix, encara s'ha de tractar l'àmplia majoria de trajectòries que passen «a través» de les interseccions sense fer un gir. En aquesta fase s'identifiquen les trajectòries no tractades i s'ajusten els enllaços ja existents a aquestes dades. En aquesta fase no es creen ni nous enllaços ni noves interseccions. A l'**Algorisme 2.4** trobareu el pseudocodi d'aquesta part.

El primer pas d'aquesta fase és ordenar decreixentment els enllaços segons la seva longitud. Seguidament, per cada enllaç es troben totes les trajectòries que intersequen a l'enllaç engrandit. Per tal d'engrandir l'enllaç s'utilitza la distància màxima que

hi havia en els trossos originals. De les trajectòries que intersequen l'enllaç engrandit només es consideren els trossos que passen per l'interior (vegeu la Figura 2.5) i, a més, per tal d'evitar trajectòries que travessen l'enllaç però no són angularment similars tinguin un efecte en aquest, hi ha un llindar angular definit pel paràmetre *angle trajectòries*.

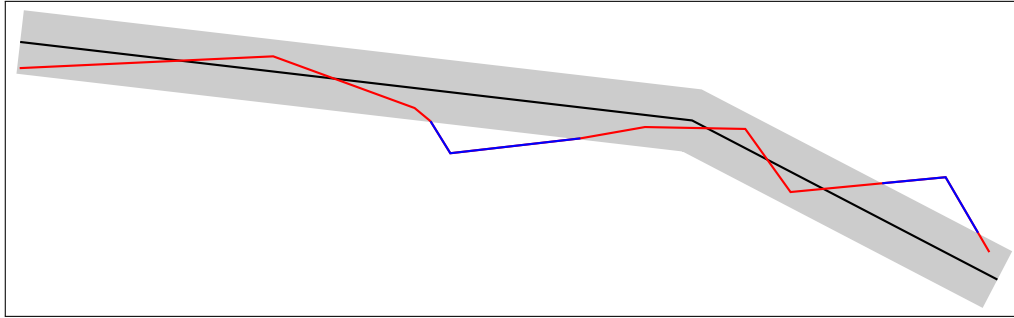


Figura 2.5: En negre l'enllaç, en gris l'enllaç engrandit i en vermell i blau la trajectòria que s'està tractant. En aquest cas, només els trossos vermells es consideraran ja que els blaus estan fora de l'enllaç engrandit.

Un cop identificats els nous trossos per cada enllaç, es procedeix a corregir l'enllaç mitjançant una agrupació similar a la fase anterior, és a dir, mitjançant l'algorisme d'escombrat (donant més pes a l'enllaç original segons el nombre de trajectòries que l'han originat).

Eliminació d'interseccions triangulars

Malgrat ja s'hagi reconstruït un mapa, aquesta fase té com a objectiu millorar-ne la qualitat mitjançant un procediment heurístic. Concretament, el fenomen que es vol combatre és la creació d'interseccions addicionals que succeeixen en corbes llargues on es detecten girs al principi i al final que acaben esdevenint interseccions separades. Aquestes interseccions extres donen peu a drecceres en el gir en forma de triangle, d'aquí el seu nom d'*intersecció triangular* (a la Figura 2.6 se'n mostra un exemple).

L'heurística té en compte la quantitat de trajectòries que han creat els enllaços del triangle. Concretament, siguin l_1 , l_2 i l_3 tres enllaços que formen un triangle, mira si un dels tres enllaços té relativament moltes menys trajectòries que els altres dos enllaços, i en cas que així sigui, l'elimina.

Algorisme 2.4 Correcció dels enllaços.

```

1: function CORREGIRENLLAÇOS( $T, L, \text{angle traj.}$ )
2:    $L \leftarrow \text{SORTDESC}(L, \text{length})$   $\triangleright$  ordena els enllaços per longitud
3:   for all  $l \in L$  do  $\triangleright$  per cada enllaç troba trajectòries properes
4:      $l_G \leftarrow \text{ENGRANDEIX}(l, l.\text{maxdist})$ 
5:      $CLS \leftarrow \text{CERCAINTERSECCIONS}(T, l_G, \text{angle traj.})$ 
6:     for all  $cl \in CLS$  do  $\triangleright$  per cada trajectòria que passa prop de  $l$ 
7:       if  $\text{INTERIOR}(l_G, cl)$  then  $\triangleright$  tota la trajectòria està continguda en  $l_G$ 
8:          $l \leftarrow \text{ESCOMBRAT}(l, cl)$ 
9:       else
10:         $cl_{in}, cl_{out} \leftarrow \text{SEPARA}(l, cl)$ 
11:         $CLS.\text{afegir}(cl_{out})$   $\triangleright$  afegir el tros encara no tractat al conjunt de no
        tractats
12:         $l \leftarrow \text{ESCOMBRAT}(l, cl_{in})$   $\triangleright$  tracta  $cl_{in}$ 
13:      end if
14:    end for
15:  end for
16: end function

```

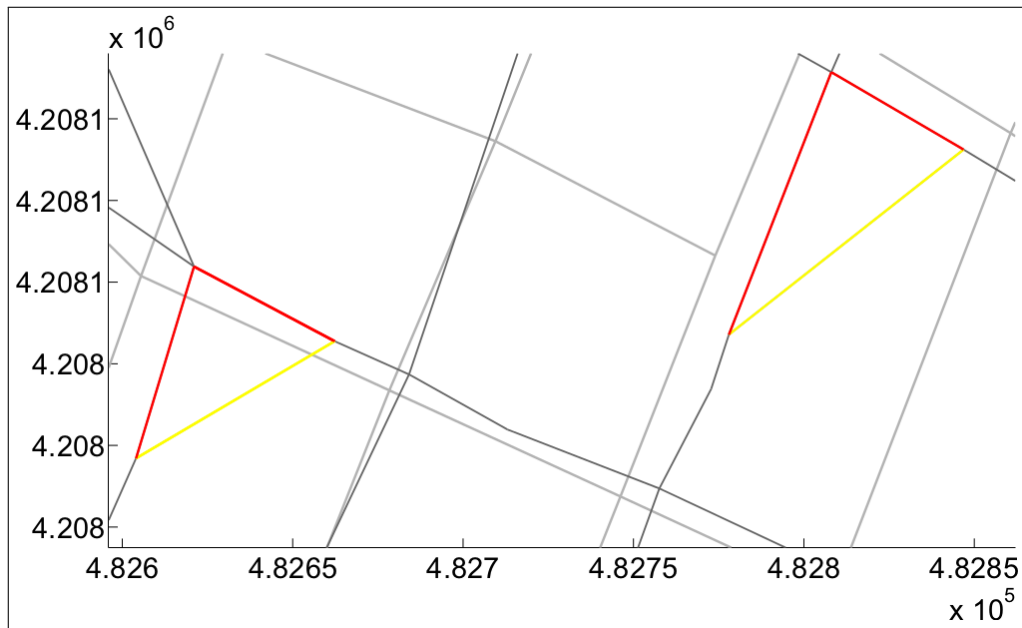


Figura 2.6: Intersecció triangular. Les arestes grogues són eliminades per comparació de la seva densitat amb la de les arestes vermelles. Imatge extreta de Karagiorgou i Pfofer [31].

2.2 Mesures de la qualitat

En la bibliografia de teoria de grafs s'han proposat diverses distàncies per comparar dos grafs abstractes, p. e. el subgraf isomorf més gran. Tanmateix, aquestes mesures no tenen en compte factors geomètrics de les immersions dels grafs al pla i, per tant, no capturen les propietats de semblança que es cerquen quan es comparen un mapa generat amb el *ground-truth* corresponent.

Per a la comparació de mapes, Ahmed et al. [5] classifica les mesures segons si es basen en una distància entre *conjunts de punts* o en una distància entre *conjunts de camins*.

Les *mesures basades en distàncies entre conjunts de punts* consideren un subconjunt del conjunt (infinit) de punts pertanyents als vèrtexs i a les arestes del graf. A partir d'aquests (sub)conjunts de punts obtinguts dels dos mapes a comparar, en calculen una distància com la distància Hausdorff (dirigida o no) [7] o les proposades per Biagioni i Eriksson [11] i Liu et al. [33]. Aquestes mesures tenen l'inconvenient que no reflecteixen l'estructura topològica dels grafs i, per tant, no en comparen propietats com la connectivitat.

Les *mesures basades en distàncies entre camins*, en canvi, consideren un seguit de camins de cada graf i els comparen. Aquestes mesures reflecteixen l'estructura topològica dels grafs i en són un exemple els proposats per Karagiorgou i Pfoser [31] i Mondzsch i Sester [34] que utilitzen els camins més curts entre diferents punts del graf o el descrit per Ahmed, Hickmann i Wenk [3] que comparen camins de longitud fixada, ajudant-se de tècniques de *map-matching*.

Cal fer esment específic que les mesures de comparació de mapes no sempre treballen sota la condició que el mapa generat i el *ground-truth* cobreixin el mateix terreny, sinó que sovint és interessant trobar la comparació del mapa generat amb un subgraf del *ground-truth*, ja que és habitual que aquest darrer cobreixi més terreny. Les tres mesures que es descriuen a continuació estan pensades per ser adequades sota aquesta condició. No obstant, algunes mesures com les proposades per Biagioni i Eriksson [11] i Liu et al. [33] requereix que ambdós mapes siguin suficientment semblants i que no n'hi hagi un que sigui un superconjunt de l'altre.

2.2.1 Distància Hausdorff dirigida

La distància Hausdorff dirigida [7] entre dos conjunts M i T es descriu com:

$$\vec{d}_H(M, T) = \max_{m \in M} \min_{t \in T} d(m, t),$$

on $d(m, t)$ acostuma a ser la distància Euclidiana quan M i T són conjunts de punts. Intuïtivament, aquesta distància agafa per cada punt $m \in M$ el seu punt més proper en T (en termes de la funció $d(m, t)$) i de totes aquestes distàncies, agafa la més gran. Quan es vol comparar un mapa generat amb el *ground-truth*, normalment s'agafen com a conjunt M els vèrtexs del mapa generat i com a T els vèrtexs del *ground-truth* o, si s'apliquen tècniques de *map-matching*, T pot arribar a ser el *ground-truth* sencer.

2.2.2 Distància basada en camins

Sigui π_M el conjunt tots els camins del mapa generat i π_T el de tots els camins del *ground-truth*. La distància basada en camins es descriu com la distància Hausdorff dirigida entre π_M i π_T agafant com a mesura de distància dels elements la distància Fréchet. Més concretament, la distància basada en camins és:

$$\vec{d}_{PB}(\pi_M, \pi_T) = \max_{p_M \in \pi_M} \min_{p_T \in \pi_T} \delta_F(p_M, p_T),$$

on

$$\delta_F(p_M, p_T) = \inf_{\alpha, \beta: [0,1] \rightarrow [0,1]} \max_{t \in [0,1]} d(p_M(\alpha(t)), p_T(\beta(t))),$$

sent d la distància Euclidiana i α i β parametrizacions monòtones creixents (a la [Secció 2.1.2](#) s'ha explicat en més detall la distància Fréchet). Idealment, π_M i π_T haurien de ser els conjunts de tots els camins en M i T , respectivament, però, malauradament, aquests tenen mida exponencial. Ahmed, Hickmann i Wenk [3] demostren com $\vec{d}_{PB}(\Pi_M, \Pi_T)$ pot ser aproximat per $\vec{d}_{PB}(\Pi_M^3, \Pi_T)$, on Π_M i Π_T són els conjunts de tots els camins del mapa generat i el *ground-truth*, respectivament, i Π_M^3 és el conjunt de tots els camins de longitud 3 del mapa generat (sense considerar vèrtexs de grau 2). La distància $\vec{d}_{PB}(\Pi_M^3, \Pi_T)$ pot ser calculada en temps polinòmic respecte la mida dels grafs aplicant tècniques de *map-matching* a T per cada camí de Π_M^3 .

2.2.3 Distància de camins més curts

La mesura de qualitat descrita per Karagiorgou i Pfoser [31] es calcula en dos passos. En un primer pas se suprimeix la geometria sobrant del *ground-truth* i, seguidament, en un segon pas es computen un nombre massiu de camins curts en el mapa generat i en el *ground-truth* i es comparen per la distància Fréchet i la distància mitjana vertical.

La poda del *ground-truth* està basada en un principi molt senzill. La idea consisteix en engrandir en 50 m les arestes del *ground-truth*, de manera que aquelles arestes engrandides que no intersequen amb cap aresta del mapa generat són eliminades. D'aquesta manera, les zones del *ground-truth* que no siguin cobertes pel mapa generat quedaran suprimides o, si més no, desconnectades i, per tant, no formaran part dels camins més curts.

La segona fase de l'algorisme consisteix en agafar parelles de punts triats a l'atzar del mapa generat. Per cada parella de punts, es busca la seva parella corresponent al *ground-truth* agafant-ne el punt més proper i dins una finestra de 100 m, si existeix. Un cop trobades ambdues parelles, es computen els camins mínims que connecten cada parella al mapa respectiu mitjançant l'algorisme de Dijkstra i es comparen a raó de la distància Fréchet discreta i la distància mitjana vertical. Aquest procediment es repeteix per un elevat nombre de parelles de punts (500 en l'article original).

3 Conjunts de dades

3.1 Preàmbul

S’han utilitzat quatre conjunts de dades per fer els experiments: *Delta del Llobregat*, *Aiguamolls Baix Empordà*, *Turó de l’home* (situat al parc natural del Montseny) i *Garraf*. Cada conjunt inclou un seguit de trajectòries de senderisme en format GPX [26] extretes de Wikiloc [37] i un mapa de la zona (*ground-truth*) obtingut de OpenStreetMap [22], ampliat lleugerament per cobrir més terreny (a la [Secció 3.4.1](#) s’explica com s’ha modificat cada mapa).

A més d’aquests quatre conjunts de dades, s’han analitzat els d’Ahmed et al. [5] que inclouen grans ciutats com *Atenes* (dos conjunts diferents, un de menys extensió, *petit*, i un que cobreix més zona, *gran*), *Berlín* i *Chicago* (tots quatre disponibles a mapconstruction.org), per tal de poder-los comparar.

Dels quatre conjunts de dades de senderisme, *Delta del Llobregat* i *Garraf* han sigut aportats íntegrament per Enle Lin (excepte les modificacions del *ground-truth*), que els ha usat en el seu treball encara no publicat [32].

La resta del capítol ve organitzat en tres parts. A la [Secció 3.2](#) es descriu breument la topografia del terreny i quins límits s’han considerat en els diferents conjunts de dades. Tot seguit, la [Secció 3.3](#) es dedica a les trajectòries i el tractament d’elles. Finalment, a la [Secció 3.4](#) es detalla quins *ground-truth* s’han utilitzat i de quina manera s’han obtingut. Al llarg de totes les seccions s’han comparat les dades de senderisme amb els conjunts de dades de ciutat d’Ahmed et al. [5].

3.2 Terrenys

Aquests quatre conjunts de dades cobreixen bona part del tipus de senderisme propi de Catalunya. Es poden dividir segons si corresponen a un senderisme de zona plana,

com és el cas del *Delta del Llobregat* i *Aiguamolls Baix Empordà*, o bé a un senderisme a un massís (zona muntanyosa) que seria el cas del *Turó de l'home* i *Garraf*. Cal destacar que el *Turó de l'home* està cobert de boscos densos mentre que en el *Garraf* la vegetació és escassa i no gaire alta, i en el *Delta del Llobregat* i als *Aiguamolls Baix Empordà* no hi ha vegetació en els camins. Per tenir una imatge general de les zones observeu la **Figura 3.1**.



Figura 3.1: D'esquerra a dreta i de dalt a baix, trobem imatges del *Delta del Llobregat*, *Aiguamolls Baix Empordà*, *Turó de l'home* i *Garraf*. En les dues primeres imatges es veuen camps de conreu així com el riu Ter i una platja en la segona. A la tercera imatge es veu el pantà de Santa Fe amb part de la carena del turó de l'home al fons, observeu com la vegetació és densa i alta. La quarta fotografia correspon a una zona muntanyosa del Garraf, plena d'herbes i mates petites.

Fonts: (1) Usuari «Pere prlpz» de Viquipèdia. Imatge disponible [aquí](#). (2) Consorci Alba-Ter. (3) Usuari «Pere prlpz» de Viquipèdia. Imatge disponible [aquí](#). (4) Usuari «Dragoncillo» de Viquipèdia. Imatge disponible [aquí](#).

Diferències amb la ciutat Les zones de senderisme pla tenen més similituds topogràfiques amb les ciutats que les zones de senderisme muntanyós. Concretament, en les zones planes els camins són rectes i de certa amplitud i intersequen entre ells amb angles pràcticament rectes, tal i com sol passar a les ciutats. En canvi, el relleu de les zones muntanyoses fa que els camins tinguin moltes corbes, siguin estrets i connectin els uns amb els altres amb angles molt petits. A més, mentre que en el terreny pla dos camins diferents són distingibles només amb les coordenades de latitud i longitud (com en les ciutats), en les muntanyes dos camins poden ser molt similars en latitud i longitud però en canvi diferir molt a causa de la diferència d'elevació entre ells. Aquestes diferències en el terreny, que tenen una incidència en com són els mapes generats, fa d'entrada que els algorismes analitzats actuïn millor en el *Delta del Llobregat* i *Aiguamolls Baix Empordà* que en el *Turó de l'home* i el *Garraf*. Una taula comparativa dels conjunts de dades de senderisme i de ciutat la podeu trobar a [Taula 3.1](#).

Lloc	Àrea (km ²)	Característiques
CIUTAT		
<i>Atenes gran</i>	12 × 14	Carrers amples i rectes, interseccions amb angles rectes (sobretot a Chicago), bona separació entre carrers, sense elevació, sense vegetació.
<i>Atenes petit</i>	2.6 × 6	
<i>Berlín</i>	6 × 6	
<i>Chicago</i>	7 × 4.5	
PLA		
<i>Delta del Llobregat</i>	5 × 2.8	Camins força amples, sense bona separació entre camins paral·lels.
<i>Baix Empordà</i>	9.6 × 5.9	
MUNTANYA		
<i>Turó de l'home</i>	8.2 × 4.7	Camins estrets i corbats, interseccions amb angles petits, sense bona separació (en \mathbb{R}^2) entre camins, diferències d'elevació importants i un té vegetació.
<i>Garraf</i>	6.7 × 4.5	

Taula 3.1: Taula resum de l'àrea coberta per cada conjunt de dades així com de les característiques en les ciutats i les diferències en el senderisme respecte la ciutat. L'àrea està expressada com a distància est-oest per distància nord-sud.

3.2.1 Delta del Llobregat

L'àrea coberta pel joc de proves del *Delta del Llobregat* ve definida per una finestra rectangular de 2.8 km (nord-sud) per 5 km (est-oest) (vegeu la [Taula 3.2](#)). Aquesta zona es troba a nivell del mar i no presenta pendents importants. En aquest espai es troba

la desembocadura del riu Llobregat. Per tant, hi abunden canals de rec i camps de conreu i aquests, a la seva vegada, fan que els camins siguin rectes i intersequin de forma gairebé perpendicular. Cal mencionar que hi ha camins completament paral·lels que disten relativament poc entre ells (15–20 m).

	Latitud	Longitud	Nord	Est
<i>mínim (SO)</i>	41.286 377	2.082 439	4 570 954 m	423 166 m
<i>màxim (NE)</i>	41.311 783	2.141 361	4 573 724 m	428 128 m

Taula 3.2: Finestra rectangular de tall per al *Delta del Llobregat*, definida en el sistema de coordenades WGS 84, i nord i est en la projecció UTM (zona 31T).

3.2.2 Aiguamolls Baix Empordà

L'àrea coberta pel joc de proves dels *Aiguamolls Baix Empordà* (també coneguts com a Aiguamolls del Baix Ter) ve definida per una finestra rectangular de 5.9 km (nord-sud) per 9.6 km (est-oest) (vegeu la [Taula 3.3](#)). Aquesta zona és molt similar al *Delta del Llobregat* a causa de la presència de la desembocadura del riu Ter. Com en el Delta, hi ha abundància de camps de conreu, fent que els camins siguin rectes i intersequin amb angles molt pronunciats. Novament, aquesta zona es troba a nivell de mar i no presenta gaires elevacions. Tot i les similituds, en aquesta àrea no hi ha gaires camins completament paral·lels. Cal finalment remarcar la presència d'una platja d'uns 50–70 m de sorra fora del mar.

	Latitud	Longitud	Nord	Est
<i>mínim (SO)</i>	41.988 487	3.085 602	4 648 501 m	507 090 m
<i>màxim (NE)</i>	42.041 506	3.202 025	4 654 404 m	516 720 m

Taula 3.3: Finestra rectangular de tall per als *Aiguamolls Baix Empordà*, definida en el sistema de coordenades WGS 84, i nord i est en la projecció UTM (zona 31T).

3.2.3 Turó de l'home

L'àrea coberta pel joc de proves del *Turó de l'home* ve definida per una finestra rectangular de 4.7 km (nord-sud) per 8.2 km (est-oest) (vegeu la [Taula 3.4](#)). El turó de l'home se situa dins del parc natural del Montseny i és la muntanya més alta de tot el

massís (1705 m). La topografia de la zona és la pròpia d'un senderisme de muntanyes on no es requereix l'escalada per arribar al cim. Els camins poden ser força estrets i hi ha interseccions de senders així com girs molt aguts, normalment per salvar l'altitud i fer camins amb menys pendent. A excepció de les cotes més altes, al llarg del vessant hi ha força vegetació, sobretot fagedes i avetoses entre altres arbres propis dels boscos centreeuropeus. L'àrea coberta per aquest conjunt de dades no només inclou el turó, sinó també els seus voltants amb zones emblemàtiques com el pantà de Santa Fe i l'ermita de Santa Fe.

	Latitud	Longitud	Nord	Est
<i>mínim (SO)</i>	41.752 525	2.418 509	4 622 463 m	451 656 m
<i>màxim (NE)</i>	41.795 680	2.516 904	4 627 203 m	459 863 m

Taula 3.4: Finestra rectangular de tall per al *Turó de l'home*, definida en el sistema de coordenades WGS 84, i nord i est en la projecció UTM (zona 31T).

3.2.4 Garraf

L'àrea coberta pel conjunt de dades del *Garraf* ve definida per una finestra rectangular de 4.5 km (nord-sud) per 6.7 km (est-oest) (vegeu la [Taula 3.5](#)). La zona pertany al massís del Garraf situat a la Serralada Litoral Catalana. Les alçades màximes del massís arriben als 600 m i les muntanyes són de tipus calcari amb avencs i coves. La vegetació de la zona no és molt densa i hi creixen només herbes i mates petites. Com en el *Turó de l'home*, els camins intersequen entre ells de forma força aguda i presenten moltes corbes. En la zona delimitada s'inclou alguna urbanització costanera, i d'entre elles n'hi ha una de la qual en surten la majoria de trajectòries al haver-hi allà una estació de tren.

	Latitud	Longitud	Nord	Est
<i>mínim (SO)</i>	41.243 725	1.861 402	4 566 438 m	404 594 m
<i>màxim (NE)</i>	41.285 209	1.940 710	4 570 959 m	411 296 m

Taula 3.5: Finestra rectangular de tall per al *Garraf*, definida en el sistema de coordenades WGS 84, i nord i est en la projecció UTM (zona 31T).

3.3 Trajectòries

Les trajectòries dels diferents conjunts de dades s'han obtingut de Wikiloc [37]. S'ha optat per agafar les dades d'aquest portal per la seva gran quantitat de trajectòries disponibles en zones de senderisme, a diferència d'altres fonts com són les trajectòries distribuïdes per OpenStreetMap. Les trajectòries de Wikiloc provenen dels seus usuaris que les han capturat de manera individual. Com a conseqüència, les dades són molt heterogènies i, a fi de reduir aquesta heterogeneïtat, només s'han considerat aquelles trajectòries sota la categoria de senderisme o caminada (a la [Taula 3.6](#) trobareu les desviacions estàndard relatives en velocitat i temps de mostreig). Finalment, una imatge global de les trajectòries considerades la podeu trobar a la [Figura 3.2](#).

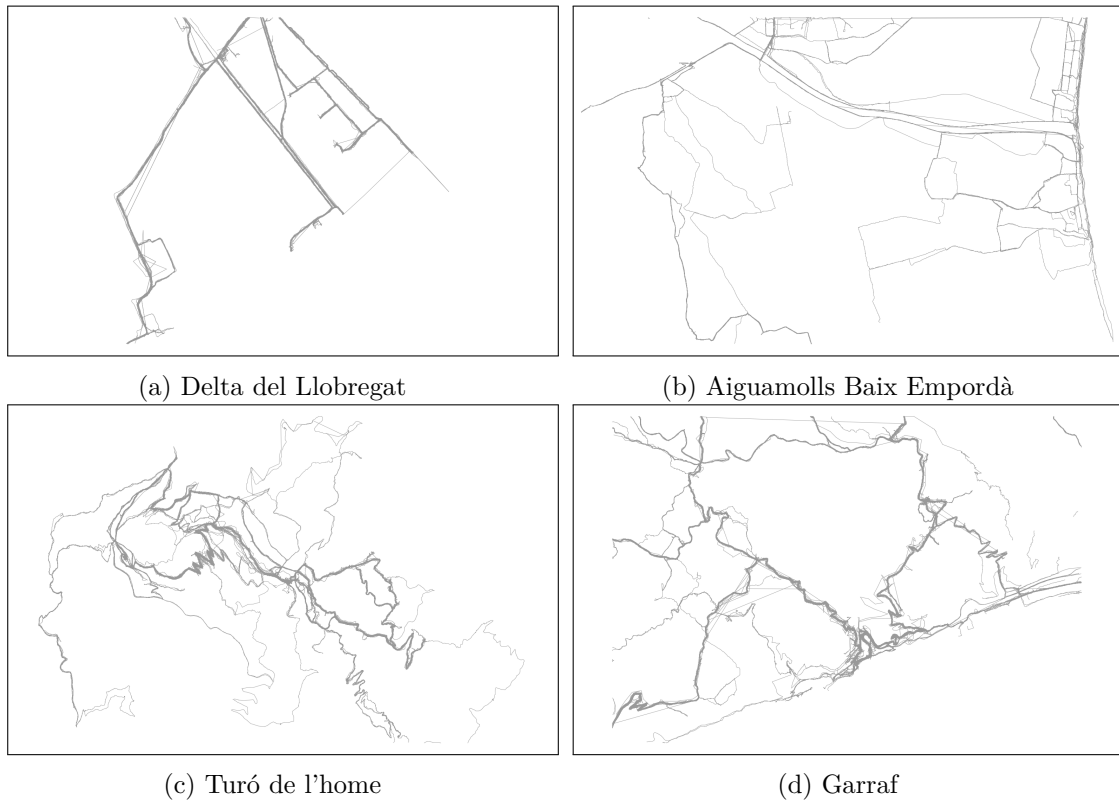


Figura 3.2: Visió global de les trajectòries dels quatre conjunts de dades.

Diferències amb la ciutat Tal i com està recollit a la [Taula 3.6](#), la complexitat (en termes del nombre total de nodes) de les trajectòries dels nostres conjunts de dades és similar a la dels de ciutat. Com era de suposar, les magnituds de distància recorreguda i

velocitat mitjana són notòriament més altes en els conjunts de dades de ciutat, ja que les dades han sigut preses des de vehicles. Els temps de mostreig de punts són força alts en *Atenes* i *Berlín* (30 s i 40.41 s) en comparació a la resta de conjunts on la mitjana ronda 2.88–10.15 s entre mostres. Per acabar, convé recalcar que en les zones del *Garraf* i del *Turó de l'home* les trajectòries tenen tendència a tenir més error ja que la recepció GPS pot veure's bloquejada o distorsionada pel relleu muntanyós així com pels boscos [41] en el cas del *Turó de l'home*.

LLoc	Traj.	Nodes	L. [km]	V. [km/h]	T. m. [s]
CIUTAT					
<i>Atenes gran</i>	120	72 439	13 432	20.25 (66.73%)	30.00 (0.00%)
<i>Atenes petit</i>	129	2840	449	18.87 (59.39%)	30.00 (0.00%)
<i>Berlín</i>	27 189	192 223	40 603	24.15 (37.40%)	40.41 (38.70%)
<i>Chicago</i>	889	118 360	2869	33.46 (29.70%)	2.88 (42.45%)
PLA					
<i>Delta del Llobregat</i>	157	38 028	474	4.70 (40.08%)	10.15 (55.25%)
<i>Baix Empordà</i>	95	46 116	493	4.88 (26.18%)	7.29 (79.06%)
MUNTANYA					
<i>Turó de l'home</i>	101	128 181	960	3.21 (48.21%)	8.56 (77.33%)
<i>Garraf</i>	629	288 484	2864	4.26 (38.40%)	8.65 (64.23%)

Taula 3.6: Taula amb algunes dades de les trajectòries. *Traj.* és el nombre de trajectòries, *Nodes* és el total de vèrtexs en totes les trajectòries, *L.* la suma de les seves longituds, *V.* és la velocitat mitjana i entre parèntesi la desviació estàndard relativa, *T. m.* és el temps de mostreig entre punts i entre parèntesi la desviació estàndard relativa. Cal mencionar que per les velocitats i els temps de mostreig s'han eliminat el 5 % de valors més baixos i el 5 % de valors més alts per tal de filtrar *outliers*. Nota: les dades dels conjunts de dades de ciutat s'han recalculat de nou i hi ha incongruències amb Ahmed et al. [5].

3.3.1 Tractament de les trajectòries

Durant el processat de les trajectòries per canviar el seu format (vegeu la [Secció 4.1.1](#)), hem dut a terme dos altres tractaments. El primer tractament és la segmentació i eliminació de les porcions de trajectòries exteriors a les finestres definides a la [Secció 3.2](#). El segon tractament està destinat a combatre un problema en les marques de temps. Cada trajectòria defineix una seqüència de punts ordenats temporalment i on cada

un dels punts conté una marca de temps, de tal manera que dos punts consecutius haurien de tenir les seves marques de temps ordenades creixentment. Aquest fet no es dona en aproximadament un 10% de les trajectòries, en les quals es presenten salts enrere significatius que varien des de dies a anys (vegeu la [Figura 3.3](#)). A més, aquestes desviacions temporals no presenten cap patró concret ni en la magnitud dels salts ni en com aquests es defineixen. Considerant la manca d'un patró clar conjuntament amb el fet que la taxa d'incidència és prou petita, les trajectòries amb aquesta anomalia han sigut directament descartades.

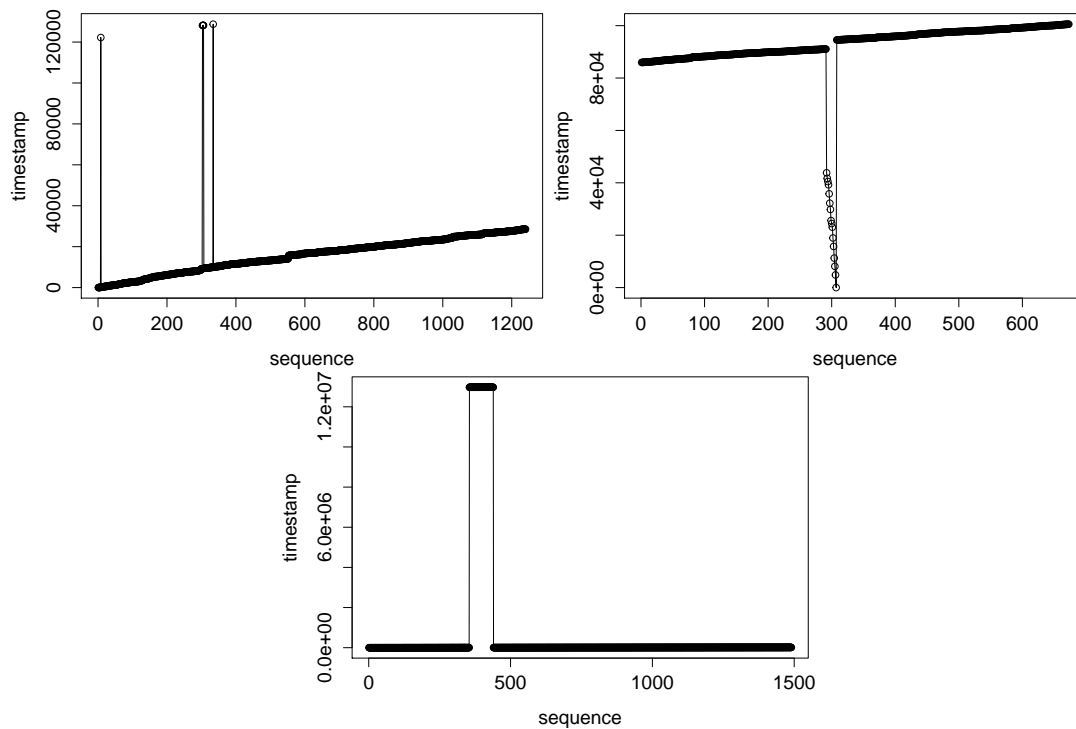
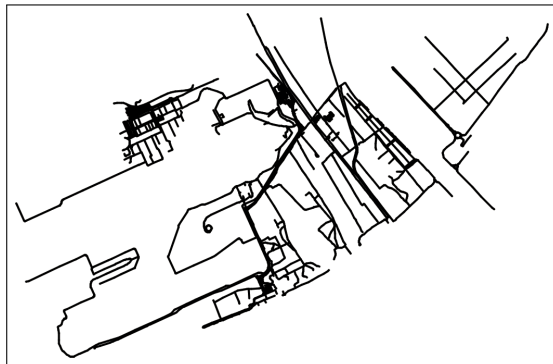


Figura 3.3: Diferents mostres d'errors en les marques del temps on en l'eix de les x hi ha l'ordre de la seqüència de punts i en l'eix de les y la marca de temps traslladada a 0. Aquestes gràfiques haurien de ser monòtones creixents. Tanmateix, en el primer gràfic s'observen pics puntuals d'unes 30 h de diferència. A la segona gràfica s'observa un interval on les marques de temps decreixen constantment fins arribar a una diferència de 22 h. Finalment, a la tercera gràfica hi ha un interval desplaçat en uns 9 anys.

3.4 *Ground-truth*

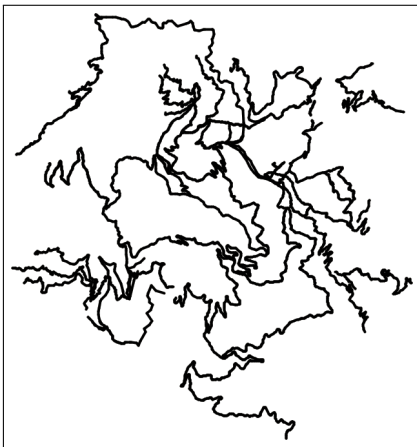
La base per als *ground-truth* dels diferents conjunts de dades han sigut extretes de OpenStreetMap [22] (en format OSM-XML [23]) ja que en molts casos, sobretot en el senderisme, és l'única font de la qual obtenir cartografia digitalitzada i de lliure distribució. No obstant, aquests mapes no són lliures d'errors ni de mancances, i això és especialment cert en les zones rurals i naturals així com poc poblades, i és per aquest motiu que hem rectificat sistemàticament els *ground-truth* tal i com es descriu a la Secció 3.4.1. Podeu veure els diferents *ground-truth* després de ser modificats a la Figura 3.4.



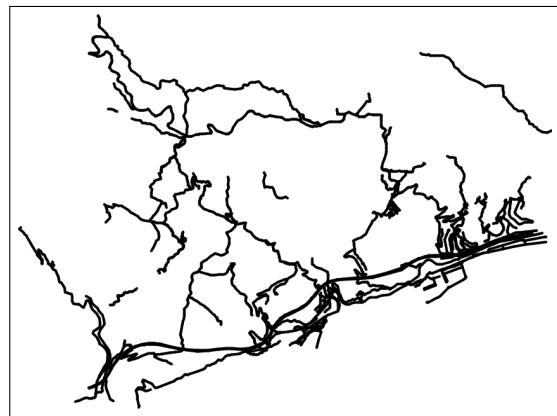
(a) Delta del Llobregat



(b) Aiguamolls Baix Empordà



(c) Turó de l'home



(d) Garraf

Figura 3.4: Visió global dels quatre *ground-truth*. Cal dir que els *ground-truth* cobreixen més àrea que la delimitada ja que també s'inclouen camins que intersequen la zona i no hi tenen els seus extrems.

Diferències amb la ciutat Deixant de banda les diferències en la geometria subjacent que volen representar els *ground-truth* de ciutat envers als de senderisme (tractades anteriorment a la Secció 3.2), la diferència més important és la qualitat dels *ground-truth* de senderisme comparats amb els de ciutat. Malgrat que les densitats del nombre de vèrtexs per l'àrea coberta així com el nombre de vèrtexs per la longitud de tot el *ground-truth* siguin similars en la muntanya i les ciutats (vegeu la Taula 3.7), aquestes mesures no reflecteixen el problema principal que presenten els de muntanya: la falta de camins. Per tal d'alleujar aquest problema, s'han introduït diverses rectificacions però encara són forces els camins del *Turo de l'home* i del *Garraf* que no surten representats als seus respectius *ground-truth*.

Lloc	Vèrtexs	Arestes	Long. (km)	Densitat vèrtexs	
				u/km	u/km ²
CIUTAT					
<i>Atenes gran</i>	32 212	39 699	2000	16.11	191.74
<i>Atenes petit</i>	2694	3436	193	13.96	172.69
<i>Berlín</i>	5894	6839	360	16.37	163.72
<i>Chicago</i>	9429	11 801	605	15.59	299.33
PLA					
<i>Delta del Llobregat</i>	3324	3454	110	30.22	237.43
<i>Baix Empordà</i>	3618	3834	241	15.01	63.88
MUNTANYA					
<i>Turó de l'home</i>	6919	6931	137	50.50	179.53
<i>Garraf</i>	4358	4390	119	36.62	144.54

Taula 3.7: Taula amb algunes dades dels *ground-truth*. La longitud (long.) fa referència a la suma de longituds de totes les arestes. La densitat de vèrtexs està mesurada pel nombre de vèrtexs dividit per la longitud del *ground-truth* (u/km) i pel nombre de vèrtexs dividit per l'àrea coberta (u/km²) definida a la Taula 3.1.

3.4.1 Tractament dels *ground-truth*

Les dades d'OpenStreetMap contenen bàsicament dos tipus d'error, d'una banda hi ha camins desplaçats de la realitat, i de l'altra hi ha camins que no hi surten representats. Per al propòsit d'aquest treball, el primer tipus d'error no suposa gaires problemes, ja que la seva ocurrència és baixa i els errors estan fitats. Tanmateix, el segon tipus d'error presenta

un problema important, ja que és origen d'*outliers* en les mesures de qualitat emprades i es donen amb força freqüència en el *Turó de l'home* i el *Garraf*.

Amb el propòsit d'alleujar la manca de camins, s'han millorat manualment (amb el programa JOSM [29]) els *ground-truth* per tal que cobrissin més terreny. Per a aquest fi, s'ha utilitzat la imatgeria via satèl·lit de Bing [17], que ha sigut suficient per al *Delta del Llobregat* i els *Aiguamolls Baix Empordà*, gràcies a l'escassetat de vegetació i el contrast dels camins en les imatges. Per al *Turó de l'home* i el *Garraf* s'han usat addicionalment mapes digitals de la diputació de Barcelona dels seus respectius parcs naturals [9, 10] així com trajectòries GPS d'OpenStreetMap, amb l'objectiu de superar la dificultat de distingir els camins amb només la imatgeria per satèl·lit. Convé recalcar que les rectificacions es basen en dades completament alienes a les trajectòries que després els algorismes utilitzaran. Aquest fet és imprescindible per tal de garantir la validesa final dels resultats obtinguts.

Les diferents mesures de qualitat que utilitzem eliminen o ignoren les porcions del *ground-truth* sense correspondència en el mapa generat. Conseqüentment, només les zones per on passen trajectòries i no hi ha el corresponent camí en el *ground-truth* són zones problemàtiques. Tenint aquesta idea present, les trajectòries s'han utilitzat exclusivament per identificar aquestes zones problemàtiques i així reduir al màxim el nombre de canvis realitzats en els *ground-truth*.

Alguns exemples de tipus de rectificacions estan il·lustrats a la [Figura 3.5](#) i a la [Figura 3.6](#), mentre que a la [Figura 3.7](#) hi ha un exemple de zona que no s'ha pogut completar. A la [Figura 3.8](#) es pot veure la diferència dels camins coberts tant en les dades originals com després de rectificar-les. Per veure el nombre de canvis vegeu la [Taula 3.8](#), on hi ha un resum de les diferències en la complexitat dels *ground-truth* originals respecte els rectificats.

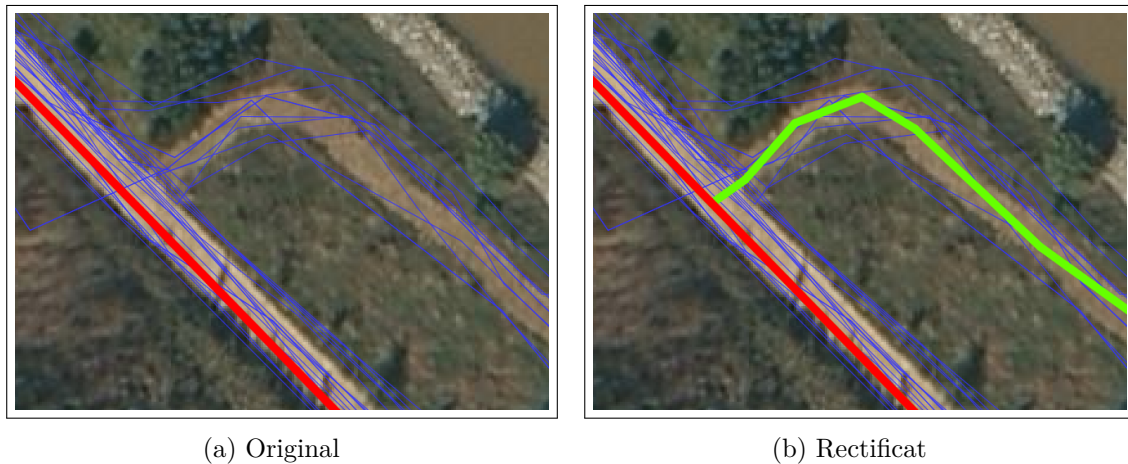


Figura 3.5: Rectificació en el *Delta del Llobregat*. En vermell el *ground-truth* original, en verd la rectificació realitzada i en blau les trajectòries del conjunt de dades. Cal recordar que aquí les trajectòries només serveixen per identificar les zones sense camins i no s'han usat per fer la rectificació. En aquesta zona, així com en els *Aiguamolls Baix Empordà*, amb la imatge via satèl·lit n'hi ha hagut prou per rectificar el *ground-truth* a causa de la bona visibilitat dels camins, tal i com es pot observar en la imatge. La imatge via satèl·lit prové de Bing Maps.

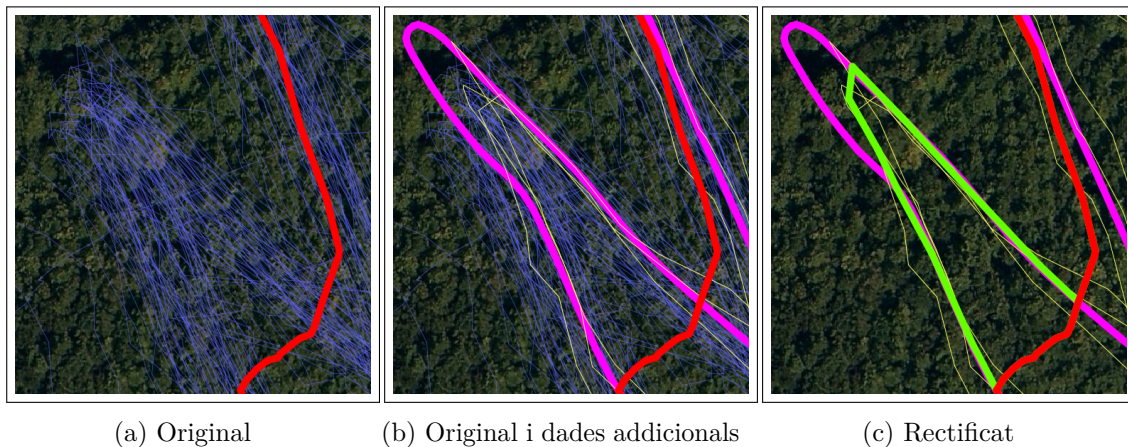


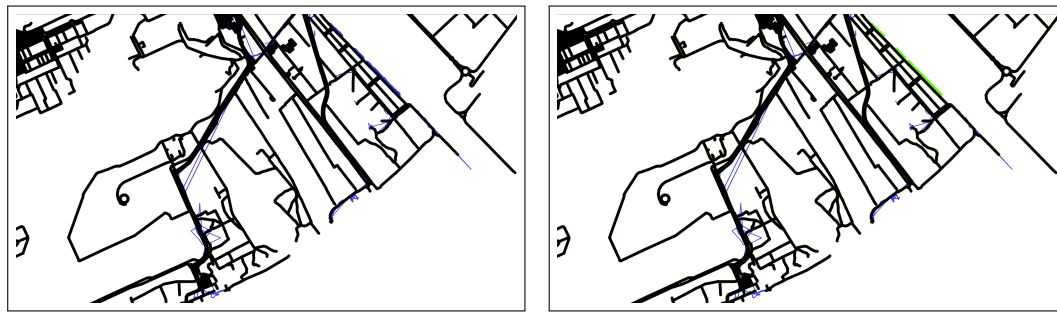
Figura 3.6: Rectificació en el *Turó de l'home*. En vermell el *ground-truth* original, en verd la rectificació realitzada, en blau les trajectòries, en morat el mapa de la diputació de Barcelona i en groc les traces GPS d'OpenStreetMap. Tal com es pot observar a la subfigura (a), la imatge per satèl·lit no és suficient per identificar el camí ja que la vegetació el cobreix. Tanmateix, agafant les dades addicionals de la diputació de Barcelona i les trajectòries d'OpenStreetMap es pot crear el camí. La imatge via satèl·lit prové de Bing Maps.



Figura 3.7: Zona del *Turó de l'home* on per la densitat de les trajectòries es pot deduir un camí. Tanmateix, ni la imatge per satèl·lit ni les dades addicionals el mostren, per tant no és afegit al *ground-truth*. La imatge via satèl·lit prové de Bing Maps.

Lloc	Vèrtexs	Arestes	Long. (km)
DELTA DEL LLOBREGAT			
<i>Original</i>	3250	3377	109
<i>Rectificat</i>	3324	3454	110
<i>Increment</i>	74	77	1
BAIX EMPORDÀ			
<i>Original</i>	3421	3629	237
<i>Rectificat</i>	3618	3834	241
<i>Increment</i>	197	205	4
TURÓ DE L'HOME			
<i>Original</i>	6042	6053	125
<i>Rectificat</i>	6919	6931	137
<i>Increment</i>	877	878	12
GARRAF			
<i>Original</i>	3737	3766	110
<i>Rectificat</i>	4358	4390	119
<i>Increment</i>	621	624	9

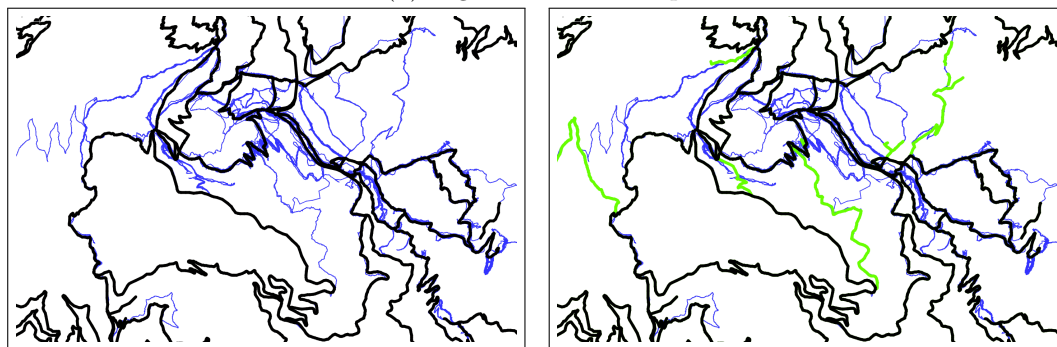
Taula 3.8: Taula amb les diferències entre els *ground-truth* originals i els rectificats.



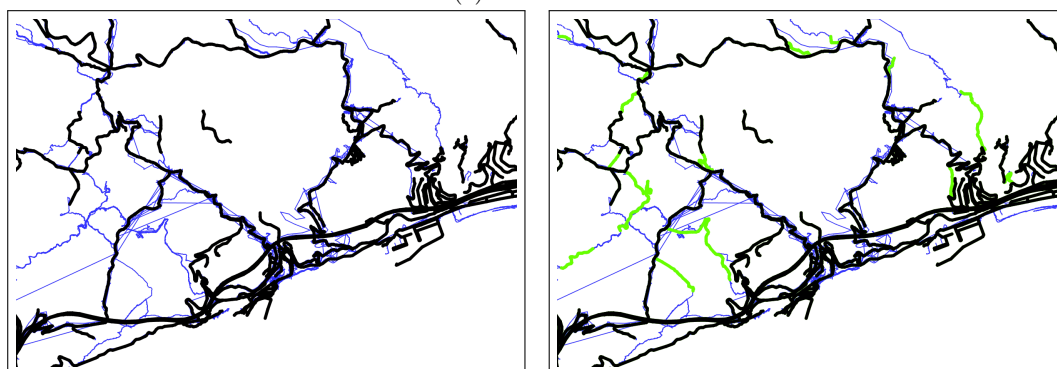
(a) Delta del Llobregat



(b) Aiguamolls Baix Empordà



(c) Turó de l'home



(d) Garraf

Figura 3.8: En negre el *ground-truth* original, en verd les rectificacions i en blau les trajectòries. Aquestes imatges il·lustren tant les rectificacions com els camins no coberts pel *ground-truth* definitiu.

4 Experimentació

Per tal de fer una valoració de la qualitat dels algorismes d'Ahmed i Wenk [4] i de Karagiorgou i Pfoer [31] en la generació de mapes a partir de trajectòries de senderisme, s'ha fet una experimentació estructurada en dues parts. La primera part ha consistit en l'anàlisi dels paràmetres i en el seu ajustament mitjançant inspecció visual en els diferents conjunts de dades. Un cop fixats els paràmetres adequats, la segona part de l'experimentació ha consistit en provar algunes de les mesures qualitatives descrites per Ahmed et al. [5].

El codi original dels algorismes de Ahmed i Wenk [4] i Karagiorgou i Pfoer [31] està escrit en Java i MATLAB, respectivament, i s'ha obtingut de mapconstruction.org. Del mateix portal també s'han obtingut els codis originals per al càlcul de les mesures de qualitat, on la distància Hausdorff dirigida i la distància basada en camins [3] estan implementats en Java, mentre que la distància basada en camins més curts [31] es troba disponible en MATLAB.

La resta del capítol ve organitzat en cinc parts. A continuació, a la **Secció 4.1** es presenta l'entorn de treball en el qual s'han realitzat els experiments. Seguidament, a les següents tres seccions es presenten i discuteixen la parametrització dels algorismes (**Secció 4.2**), l'anàlisi per inspecció visual dels artefactes que aquests generen (**Secció 4.3**) i una valoració de l'adequació d'aquests al senderisme (**Secció 4.4**). Finalment, a la **Secció 4.5** s'avaluen les mesures de qualitat amb els diferents *ground-truth*, comparant-les amb les obtingudes per Ahmed et al. [5].

4.1 Descripció de l'entorn de treball

A partir dels conjunts de dades descrits al **Capítol 3**, dels codis originals d'Ahmed i Wenk [4] i Karagiorgou i Pfoer [31] i dels codis de les mesures de qualitat [3, 7, 31], s'ha definit un entorn de treball en el qual realitzar els experiments. Per tal de coordinar aquests tres components, s'han creat en Python diferents programes intermediaris

destinats a fer el canvi de format de les dades i a l'execució massiva i en paral·lel dels algorismes i de les mesures de qualitat. L'entorn de treball definitiu queda descomposat en dues parts. La primera part correspon a la generació de mapes amb els paràmetres adequats (vegeu la Figura 4.1) i la segona part correspon a aplicar les mesures de qualitat sobre els mapes generats definitius (vegeu la Figura 4.2).

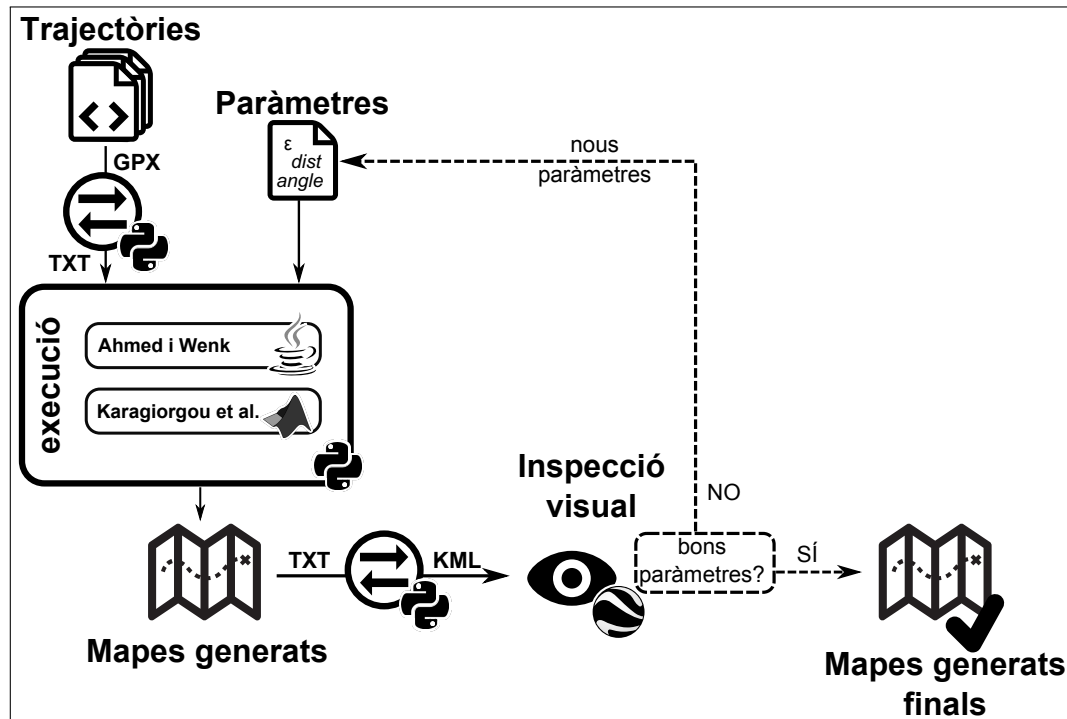


Figura 4.1: Esquema de la primera fase de l'experimentació: parametrització dels algorismes i obtenció dels mapes generats definitius.

4.1.1 Format de les dades

El format de les dades és la representació física que aquestes tenen. En entorns de treball heterogenis, és a dir, en entorns on els diferents components s'han obtingut de fonts diverses, és freqüent que els formats de dades en què cada component treballa no concordin. En aquest sentit, en el nostre entorn de treball hem distingit quatre tipus de dades (els quals s'han d'adaptar a diferents formats): *coordenades*, *trajectòries*, *mapes generats* i *ground-truth*.

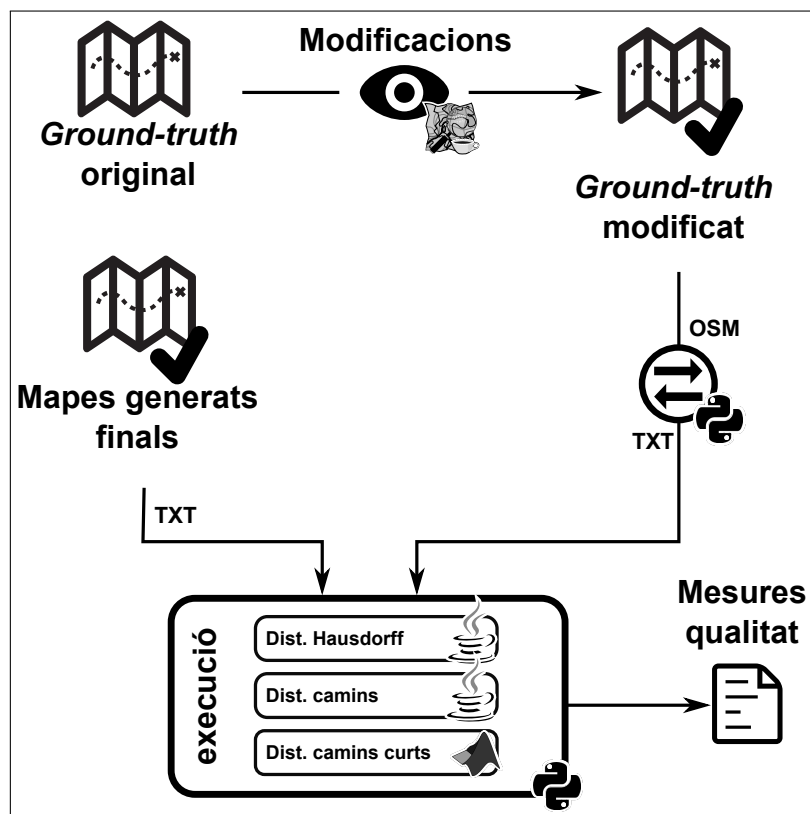


Figura 4.2: Esquema de la segona fase de l'experimentació: mesures quantitatives de qualitat.

Formats de les coordenades

Una coordenada geogràfica consisteix en una latitud i una longitud, combinats amb un sistema de referència espacial (CRS, de l'anglès *Coordinate Reference System*) [36]. En el CRS hi ha especificat quin model el·lipsoidal s'usa en l'aproximació de la superfície terrestre i, per tant, a quins punts fan referència les latituds i longituds. Cal fer esment específic al fet que la tecnologia GPS utilitza coordenades geogràfiques que tenen com a CRS l'estàndard World Geodetic System 1984 (WGS 84) [20].

Més enllà de les coordenades geogràfiques, molt sovint s'usen coordenades projectades en un pla, ja que això simplifica la geometria i, per tant, simplifica càlculs importants com el de la distància entre punts. Un dels sistemes de coordenades projectades més usat és l'Universal Transverse Mercator (UTM), el qual divideix la terra en seixanta regions que projecta de la manera més adequada, sent la zona 31T la zona que correspon

a Catalunya. Les coordenades del sistema de coordenades UTM estan expressades en metres.

Formats de les trajectòries

Les trajectòries de Wikiloc prenen format GPX [26] el qual és un esquema XML, és a dir, un fitxer XML amb una estructura predefinida. Aquest format especifica que cada punt ha de tenir les coordenades latitud i longitud en el sistema WGS 84, l'altitud com a opcional i expressada en metres sobre el nivell del mar, i la marca temporal en el format ISO 8601 (per exemple, «2011-11-27T09:37:12Z»).

Un format més senzill l'utilitzen els algorismes Ahmed i Wenk [4] i Karagiorgou i Pfofer [31], els quals treballen amb fitxers de text amb una seqüència ternes de x , y i t , on x i y corresponen a coordenades mètriques projectades i t als segons transcorreguts des d'un moment fix en el temps.

Amb la finalitat de passar de GPX a les ternes que poden llegir els algorismes, s'ha utilitzat un *parser* XML per extreure les coordenades i els temps dels fitxers GPX, i a continuació s'han projectat les coordenades a UTM (zona 31T) i expressat els temps com els segons des de l'*epoch time* (1 de gener de 1970).

Formats del mapa generat

El mapa generat pels algorismes és un graf geomètric al pla que s'emmagatzema en dos fitxers, un per vèrtexs i l'altre per arestes. El fitxer de vèrtexs està compost per ternes id , x i y , on id és un nombre identificador del vèrtex i x i y són les seves coordenades en el mateix sistema de coordenades que les trajectòries d'entrada (en els nostres conjunts de dades, UTM zona 31T). El fitxer d'arestes consisteix en ternes id , $id1$ i $id2$, on id és un nombre identificador de l'aresta i $id1$ i $id2$ corresponen als dos identificadors dels vèrtexs units per l'aresta. En l'àmbit d'aquest projecte, el graf es tracta com a no dirigit i, per tant, cada aresta en representa dues: $(id1, id2)$ i $(id2, id1)$.

El format dels fitxers de vèrtexs i arestes és el mateix que utilitzen els codis per a la mesura quantitativa de qualitat dels mapes generats. No obstant, aquest format no és pràctic per a eines de visualització amb imatgeria via satèl·lit. A fi de fer l'anàlisi visual amb Google Earth, s'han convertit els grafs geomètric a format KML (Keyhole Markup Language) [25], que és un esquema XML destinat a expressar informació geomètrica.

El format KML permet emmagatzemar diferents geometries com punts, polígons, línies poligonals, etcètera. No obstant, no té integrat cap forma de representar grafs geomètrics. Com a conseqüència, la transformació del graf geomètric a KML passa per un pas intermedi on s'obté un conjunt de camins disjunts del graf tals que continguin totes les arestes d'aquest. Finalment, el KML generat conté una línia poligonal per cada un dels camins d'aquest conjunt de camins del graf.

Formats *ground-truth*

Tota la geometria d'OpenStreetMap està expressada per quatre elements bàsics: *nodes*, *camins*, *relacions* i *etiquetes*. El format amb el qual s'ha treballat és OSM-XML [23], el qual reproduïx aquests quatre tipus de dades en un fitxer XML.

Per altra banda, el format que esperen els codis de les mesures de qualitat és el mateix que el del mapa generat: un graf geomètric descrit per dos fitxers, un per als vèrtexs i l'altre per a les arestes.

Per tal d'obtenir els *ground-truth* en el format dels dos fitxers, s'ha utilitzat un *parser* específic per al format OSM-XML. Un cop llegida tota la geometria de la zona en qüestió, s'han extret tots els camins (juntament amb els seus respectius nodes) que contenen la paraula clau *highway* a les seves etiquetes, paraula clau que identifica qualsevol tipus de carretera, carrer o camí. Gràcies a què els nodes que representen una intersecció apareixen per duplicat en tots els camins que hi incideixen, la geometria extreta és, en efecte, el *ground-truth* en forma de graf geomètric. Cal mencionar que els *nodes* d'OSM (l'únic element georeferenciat en OSM) tenen com a sistema de coordenades el WGS 84 i s'han hagut de projectar a UTM (zona 31T) atès que els mapes generats han sigut elaborats en aquest sistema de coordenades.

4.2 Parametrització dels algorismes

Els algorismes d'Ahmed i Wenk [4] i Karagiorgou i Pfofer [31] són parametritzats, és a dir, tenen un seguit de paràmetres que s'han d'adaptar segons les dades (vegeu el **Capítol 2**). Cal recordar que Ahmed i Wenk [4] usen un sol paràmetre espacial (ε), la idea del qual és ajuntar les trajectòries que mostregen el mateix camí si són més pròximes entre elles que ε . Karagiorgou i Pfofer [31], en canvi, empren 6 paràmetres: un d'angular i un de velocitat per identificar girs, un de temporal per filtrar salts en les trajectòries, un

d'espacial per compactar girs en agrupacions de girs, un altre d'espacial per compactar aquestes agrupacions de girs en interseccions i, finalment, un d'angular per compactar trajectòries en arestes. A la [Taula 4.1](#) hi ha un resum dels paràmetres de cada algorisme i els valors que prenen en el seu article original respectiu.

Paràmetres	Valors
BERLÍN – AHMED I WENK [4]	
ε	82.5 m
ATENES – KARAGIORGOU I PFOSER [31]	
<i>angle gir</i>	15°
<i>màx. velocitat gir</i>	40 km/h
<i>dist. girs</i>	50 m
<i>dist. inter.</i>	25 m
<i>angle trajectòries</i>	45°
<i>màx. temps mostreig</i>	35 s

Taula 4.1: Paràmetres originals dels algorismes i els diferents valors que prenen en els seus articles respectius.

4.2.1 Identificació i extracció dels paràmetres

Els codis originals d'Ahmed i Wenk [4] i Karagiorgou i Pfofer [31] han sigut implementats de manera que els seus paràmetres estan codificats en el seu interior (*hard-coded*). A fi de fer la connexió entre el nostre entorn i els codis dels algorismes, aquests darrers s'han hagut de modificar de manera poguessin ser parametrizats per un programa extern a l'hora d'invocar-los. A la [Taula 4.2](#) hi ha l'estructura de fitxers dels codis de cada algorisme amb alguna informació d'aquests així com quins d'ells han hagut de ser modificats per extreure'n els paràmetres.

En el codi d'Ahmed i Wenk [4] s'ha extret el paràmetre ε sense dificultats. El codi de Karagiorgou i Pfofer [31], en canvi, ha sigut més complex d'analitzar i d'extreure'n els paràmetres:

- Els paràmetres per la detecció de girs (angular i de velocitat) així com el paràmetre espacial per ajuntar girs en agrupacions de girs s'han pogut extreure directament. Cal dir que s'ha desestimat l'ús de la velocitat com a criteri vàlid d'identificació de girs, ja que aquest no té sentit en l'àmbit del senderisme.

Fitxer	# línies	# funcions
AHMED I WENK [4]		
Edge.java	69	3
IntervalComparatorEdge.java	67	1
Line.java	628	9
MapConstruction.java	804	11
Vertex.java	114	10
KARAGIORGOU I PFOSER [31]		
<i>llibries</i>	-	-
intersection_nodes_extraction.m	640	1
tracebundle.m	15 410	1

Taula 4.2: Fitxers de cada un dels codis dels algorismes juntament amb el seu nombre de línies i el nombre de funcions que s’hi defineixen. En negreta s’indiquen els fitxers que s’han modificat per tal d’extreure’n els paràmetres.

- El paràmetre espacial destinat a ajuntar les agrupacions de girs en interseccions es deriva (en el codi, però no en l’article) com la meitat del paràmetre espacial per ajuntar girs en agrupacions de girs. Aquesta derivació s’ha mantingut de manera que aquest paràmetre ja no juga un paper en la parametrització.
- El paràmetre de temps per filtrar salts en les trajectòries no es troba efectiu en el codi però, en canvi, hi ha un paràmetre espacial per filtrar salts en dos punts consecutius de les trajectòries que siguin més grans d’1 km. Per adaptar el aquest valor, s’ha dividit 1 km per la velocitat mitjana de les trajectòries del conjunt de dades *Atenes* i multiplicat per la velocitat mitjana de cada un dels nostres conjunts de dades.
- Sobre l’últim paràmetre, el paràmetre angular per compactar trajectòries en arestes, a causa de la poca llegibilitat i la poca modularitat del fitxer `tracebundle.m` no s’ha pogut extreure i, per tant, ha quedat fixat a 45° .

A la [Taula 4.3](#) hi ha tots els paràmetres a considerar dels dos algorismes i quins valors derivats o fixos prenen. D’ara en endavant, la parametrització dels algorismes fa referència a trobar valors adequats d’aquests nous paràmetres: ε per a Ahmed i Wenk [4] i *angle gir* i *dist. girs* per a Karagiorgou i Pfofer [31].

		Paràmetres	Valor fix
		<i>angle gir</i>	-
		<i>màx. velocitat gir</i>	∞
		<i>dist. girs</i>	-
		<i>dist. inter.</i>	$0.5 \cdot \text{dist. girs}$
		<i>angle trajectòries</i>	45°
		<i>màx. temps mostreig</i>	
		<i>màx. dist. entre punts</i>	$25 \cdot \text{dist. mitjana trajectòries}$

(a) *Ahmed i Wenk [4]*

(b) *Karagiorgou i Pfoser [31]*

Taula 4.3: Paràmetres finals dels algorismes i quins valors fixats o derivats tenen, si s'escau.

4.2.2 Cerca de valors per als paràmetres

Un cop identificats i extrets els paràmetres dels algorismes (especificats a la [Taula 4.3](#)), l'objectiu és trobar un conjunt de valors adequat per tal d'obtenir els millors mapes generats per a cada conjunt de dades.

A les [Taula 4.4](#) i [Taula 4.5](#) s'hi pot trobar sota quines parametritzacions s'han executat els algorismes i quins han sigut els valors definitius dels seus paràmetres. Per tal de facilitar-ne la comparació, a les mateixes taules també s'han inclòs els valors que prenen en els articles originals [\[4, 31\]](#) i en Ahmed et al. [\[5\]](#). Malgrat fora de l'àmbit d'aquest treball, també s'han inclòs els temps d'execució mitjans de cada algorisme en els diferents conjunts de dades.

Simultàniament a trobar les parametritzacions adequades, s'ha fet un estudi empíric dels efectes que tenen els paràmetres sobre els resultats. Aquest estudi no només justifica la tria final dels valors dels paràmetres, sinó que també exposa quines característiques presenten parametritzacions alternatives.

Paràmetres d'Ahmed i Wenk

Ahmed i Wenk [\[4\]](#) donen en el seu article unes restriccions sobre el paràmetre ε amb les quals garanteixen una qualitat teòrica del resultat (explicat en més detall a [Capítol 2](#)). Cal recordar que aquestes assumpcions són quatre:

1. Cada camí del mapa real té alguna porció separada per una distància al menys 3ε de qualsevol altre camí.

Paràmetres	Valors provats	T. d'exec. mitjà
ATENES GRAN ε [m]	180	-
ATENES PETIT ε [m]	90	-
BERLÍN ε [m]	82.5, 170 ^a	-
CHICAGO ε [m]	80	-
DELTA ε [m]	5, 10, 12.5, 15 , 20, 25, 30, 40	1.62 min
BAIX EMPORDÀ ε [m]	2.5, 5, 10, 20, 25 , 30, 35	8.63 min
TURÓ DE L'HOME ε [m]	30, 32.5, 35, 40 , 45, 50	18.48 min
GARRAF ε [m]	20, 25, 30, 35 , 40, 50	36.22 min

^aEl valor 82.5 prové de Ahmed i Wenk [4] mentre que el valor 170 prové de Ahmed et al. [5].

Taula 4.4: Taula amb les diferents parametritzacions de l'algorisme d'Ahmed i Wenk [4] que han sigut executades. En negreta s'indiquen els valors definitius dels paràmetres.

2. Si dos camins del mapa real s'apropen a una distància $\leq 3\varepsilon$, aleshores comparteixen un vèrtex v .
3. L'error de les trajectòries respecte el camí que mostregen és més petit que $\varepsilon/2$.
4. Cada trajectòria fa un recorregut acíclic en el mapa real de la zona.

Aquestes quatre condicions són massa estrictes i no es compleixen per a cap ε en cap dels quatre conjunts de dades. Concretament, hi ha trajectòries d'«anada i tornada» que violen la quarta condició en totes les zones. Per les altres tres condicions, aquestes depenen de ε , però no existeix cap ε que les compleixi simultàniament (vegeu la Taula 4.6).

Respecte a l'efecte d' ε en el mapa generat, aquest es pot observar de forma directa. Valors alts disminueixen la complexitat del resultat, fins al punt que camins diferents

Paràmetres	Valors provats	T. d'exec. mitjà
CIUTAT ^a		
<i>angle.</i> [°]	15	
<i>dist</i> [m]	50	-
<i>màx. vel.</i> [km/h]	40	
DELTA		
<i>angle.</i> [°]	15, 30, 40, 50 , 60	1.26 h
<i>dist</i> [m]	10 , 12.5, 15, 20, 25	
BAIX EMPORDÀ		
<i>angle.</i> [°]	10, 25, 30 , 40, 60	0.71 h
<i>dist</i> [m]	10, 15, 20, 25 , 30, 35, 40, 45, 60, 80	
TURÓ DE L'HOME		
<i>angle.</i> [°]	30, 60, 70	6.82 h
<i>dist</i> [m]	10, 20, 22.5, 30 , 35, 40	
GARRAF		
<i>angle.</i> [°]	30, 60, 70	29.05 h
<i>dist</i> [m]	20, 25, 30 , 35	

^aTots els conjunts de dades de ciutat han utilitzat la mateixa parametrització.

Taula 4.5: Taula amb les diferents parametritzacions de l'algorisme de Karagiorgou i Pfofer [31] que han sigut executades. En negreta s'indiquen els valors definitius dels paràmetres. Les execucions fetes han sigut un subconjunt del producte cartesià dels valors d'*angle* i de *dist*.

col·lapsen en un de sol. En canvi, valors baixos augmenten la complexitat del resultat, ja que l'algorisme considera més trajectòries, arribant al punt en què el resultat acaba sent un mapa distorsionat amb camins falsos provinents dels errors de les trajectòries. Tenint aquests fets en consideració, els valors d' ϵ triats s (els podeu trobar a la Taula 4.4) són un balanç entre aquests dos efectes indesitjables (vegeu la Figura 4.3).

Paràmetres de Karagiorgou i Pfofer

Dels sis paràmetres originals, només dos no han sigut fixats ni derivats d'altres paràmetres (vegeu la Taula 4.3): la distància per agrupar girs en agrupacions de girs (d'ara en endavant *dist*) i el llindar de l'angle per determinar els girs (d'ara en endavant *angle*).

Zona	$\varepsilon < \text{dist. camins}/3$	$\varepsilon > 2 \cdot \text{err. traj.}$
<i>Delta del Llobregat</i>	$\varepsilon < 3 \text{ m}$	$\varepsilon > 24 \text{ m}$
<i>Aiguamolls Baix Empordà</i>	$\varepsilon < 15 \text{ m}$	$\varepsilon > 45 \text{ m}$
<i>Turó de l'home</i>	$\varepsilon < 21 \text{ m}$	$\varepsilon > 42 \text{ m}$
<i>Garraf</i>	$\varepsilon < 4 \text{ m}$	$\varepsilon > 36 \text{ m}$

Taula 4.6: Condicions sobre ε en els diferents conjunts de dades. Es pot observar com les restriccions són incompatibles en totes les zones. Els valors s'han trobat per inspecció visual de les dades mitjançant Google Earth.

El paràmetre *angle* té una influència directa en el reconeixement de girs i, per tant, en el nombre d'interseccions que s'identifiquen. Idealment, *angle* hauria de ser suficient com perquè totes les interseccions reals, i només aquestes, tinguessin girs associats, tanmateix, no existeix tal valor d'*angle* en els nostres conjunts de dades, tal i com es menciona a la [Secció 4.2.3](#). Més enllà d'identificar corbes com a girs i, per tant, com a interseccions, la desviació angular també es veu incapaç de filtrar falsos positius deguts als errors propis de les trajectòries, fins i tot en les dades originals de l'article (*Atenes*), tal i com es mostra a la [Figura 4.4](#). A més, el nombre de falsos positius augmenta considerablement en dades amb molt soroll com són el cas del *Garraf* i el *Turó de l'home*.

Per construcció de l'algorisme, el paràmetre *dist* té una implicació difosa però important en el mapa generat. Principalment, *dist* té la finalitat d'agrupar els girs per acabar identificant interseccions. No obstant, tots els enllaços entre les interseccions depenen de la posició d'aquestes així com d'informacions addicionals extretes a l'hora de trobar-les. A causa d'aquest efecte en cadena, canviar el paràmetre *dist* no només afecta les interseccions que es troben, sinó que tota la geometria resultant del mapa generat també es veu afectada (vegeu la [Figura 4.5](#)).

Tenint present aquestes consideracions, la parametrització de l'algorisme ha consistit, per cada conjunt de dades, en fixar primer un valor d'*angle* adequat, és a dir, un valor que redueixi al màxim el nombre de girs identificats fora d'interseccions sense que, per contra, es perdin els girs que estan en una intersecció. A la [Figura 4.6](#) es mostren els efectes que té en la identificació de girs un valor d'*angle* insuficientment gran així com un d'excessiu. Cal mencionar que el temps d'execució es veu afectat fortament per aquest paràmetre, a tall d'exemple, *Turó de l'home* varia des de 5.56 h amb $\text{angle} = 70^\circ$ a 13.48 h si, en canvi, s'agafa $\text{angle} = 10^\circ$.

Un cop fixat el valor d'*angle*, s'ajusta el valor de *dist* experimentalment a causa de la manca d'un criteri clar per identificar els valors adequats d'aquest. Els valors que més

bons resultats han donat són valors pròxims a ε de l'algorisme de Ahmed i Wenk [4]. Aquest fet és explicable en quant ambdós paràmetres tenen objectius similars: volen agrupar mostres distorsionades d'un mateix punt (a causa dels errors de les trajectòries) a partir de la seva proximitat espacial.

Per acabar, cal fer esment específic de la gran diferència en el paràmetre *angle* entre els nostres conjunts de dades i els originals d'*Atenes*. Aquesta diferència de 15° (*Atenes*) a 70° (*Garraf* i *Turó de l'home*) es dona a causa de les diferents velocitats entre vehicles i senderistes. Suposant la mateixa freqüència de mostreig (1 s) i un mateix error GPS (5 m), un vehicle viatjant recte a $40 \text{ km/h} \approx 11.11 \text{ m/s}$ pot produir un desviament angular de 42° en la lectura GPS, mentre que, per contra, un senderista viatjant recte a $4 \text{ km/h} \approx 1.11 \text{ m/s}$ pot arribar a produir-ne gairebé el doble de 42° , concretament, un desviament angular de fins a 83.67° (vegeu la [Figura 4.7](#)).

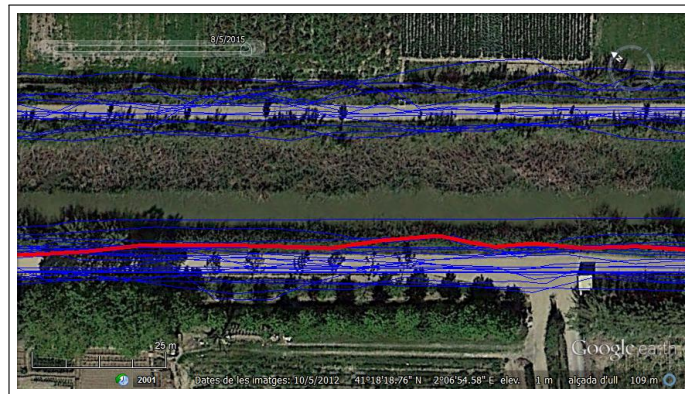
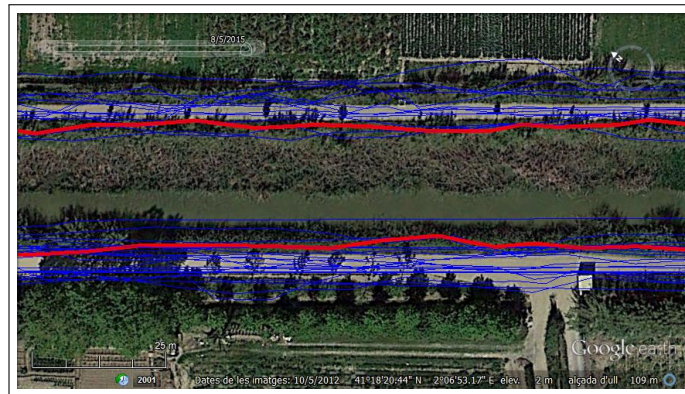
(a) ε massa baixa(b) ε massa alta(c) ε adequada

Figura 4.3: Les imatges il·lustren quin efecte causa ε en el resultat. En vermell hi ha el mapa generat i en blau les trajectòries. Imatges extremes de Google Earth en el Delta del Llobregat.

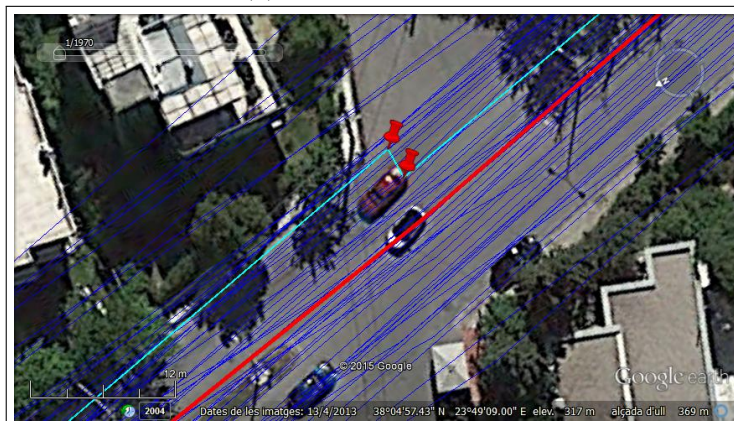
(a) *Delta del Llobregat*(b) *Atenes petit*

Figura 4.4: Exemples de trajectòries rectilínies que a causa del seu error tenen angles de gairebé 90° . En blau fosc hi ha les trajectòries, les icones vermelles indiquen els girs i en blau clar es marca la trajectòria amb l'error. Imatges extretes de Google Earth.

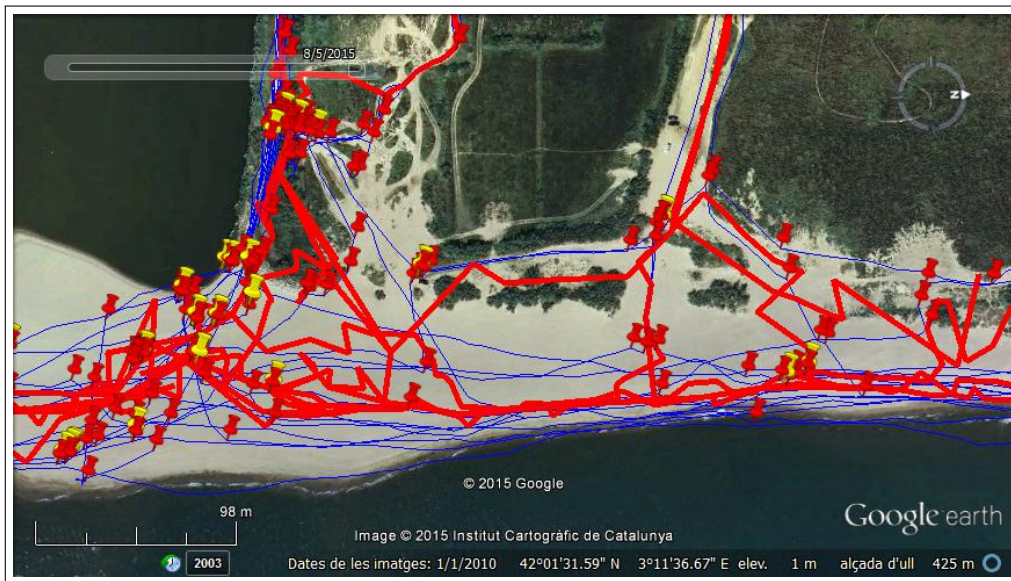
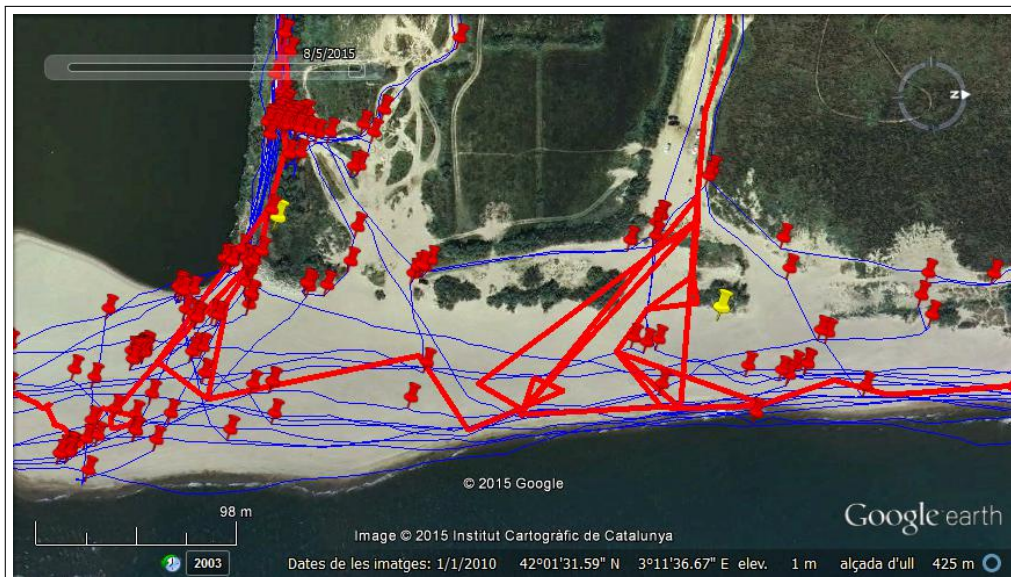
(a) valor de *dist* de 5 m(b) valor de *dist* de 100 m

Figura 4.5: Imatges que il·lustren la diferència en la geometria generada causada per canvis grans del valor del paràmetre *dist* (per a un valor d'*angle* fixat). L'efecte en cadena que un canvi en *dist* provoca en el resultat fa que no es pugui induir un patró clar de quines són les conseqüències del seu valor en el mapa generat. En vermell hi ha el mapa generat, en blau les trajectòries, les icones vermelles són els girs identificats i les icones grogues són les interseccions que es deriven dels girs identificats. Imatges extretes de Google Earth en els Aiguamolls Baix Empordà.

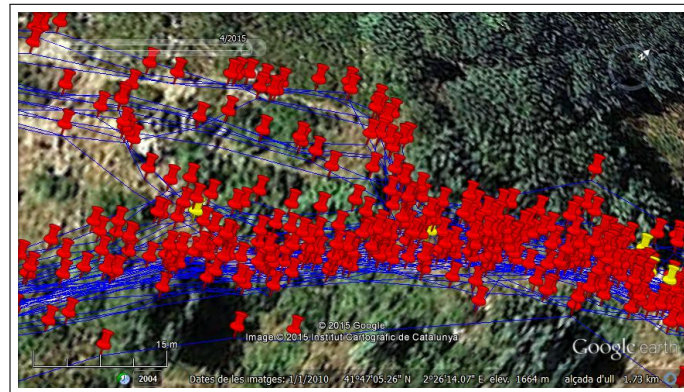
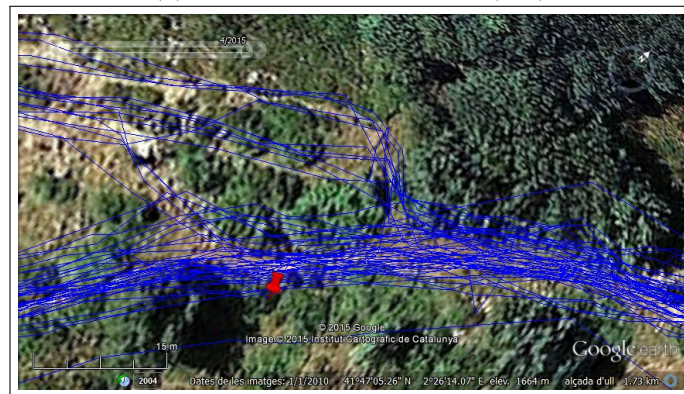
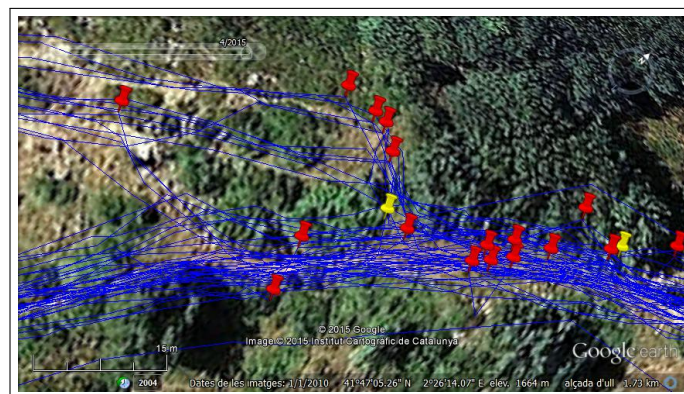
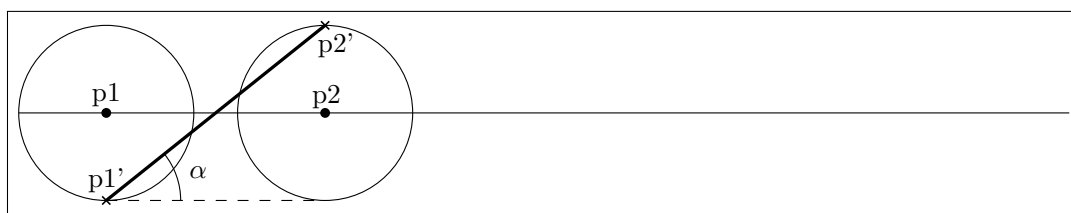
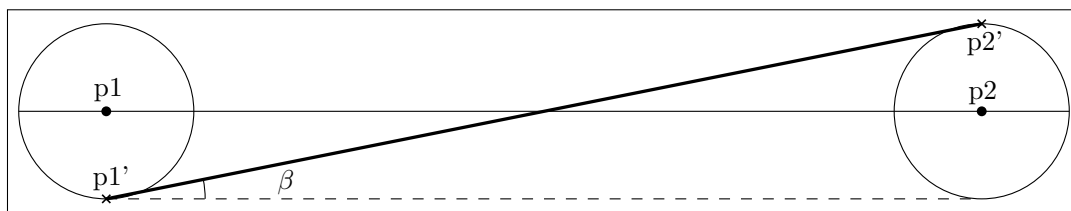
(a) valor d'angle massa baix (10°)(b) valor d'angle massa alt (85°)(c) valor d'angle adequat (60°)

Figura 4.6: Il·lustració sobre l'efecte del paràmetre *angle* en la identificació de girs. En les tres imatges *dist* ha estat fixat a 15 m. En icones vermelles els girs identificats, en icones grogues les interseccions que aquests girs generen i en blau les trajectòries. Imatges extretes de Google Earth en el *Turó de l'home*.



(a) poc recorregut entre mostres consecutives



(b) molt recorregut entre mostres consecutives

Figura 4.7: Aquests esquemes mostren que en caminar completament recte (de $p1$ a $p2$) i amb un mateix error GPS (fitat pels cercles) que desplaça $p1$ i $p2$ a $p1'$ i $p2'$ respectivament, la desviació angular és més gran quan la distància entre punts consecutius de la trajectòria és més petita.

4.2.3 Distinció angular de les interseccions

L'algorisme Karagiorgou i Pfofer [31] identifica les interseccions a partir dels girs, els quals detecta amb un criteri de reducció de velocitat més canvi angular.

A fi de fitar el llinard angular des del qual es considera un gir (que és un dels paràmetres de l'algorisme), hem analitzat en els *ground-truth* els angles que es formen en els vèrtexs intermedis (de grau 2), els quals ens interessa filtrar, i els hem comparat amb els angles que es formen en les interseccions (vèrtexs de grau > 2), que són els que busquem. En ambdós casos considerem tan sols l'angle més petit que es forma entre totes les arestes (vegeu la Figura 4.8). A la Figura 4.9 hi ha les gràfiques dels angles de cada zona i a la Taula 4.7 es recullen les dades més importants, de les quals cal destacar dos punts. Per una banda, a les ciutats el nombre d'interseccions és molt més elevat comparat amb els vèrtexs intermedis que no pas en les zones de senderisme. Per l'altra banda, cal remarcar que tret de la proporció vèrtexs intermedis—interseccions, la resta de dades són molt semblants entre ciutat i senderisme.

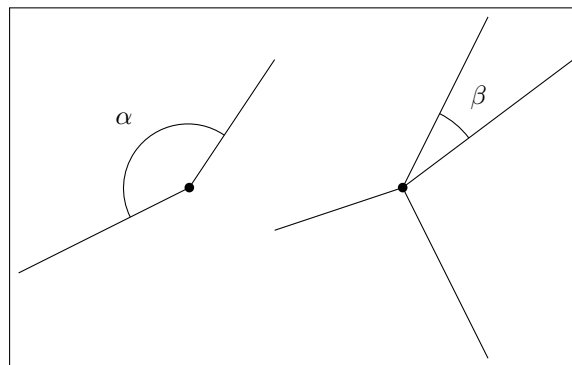


Figura 4.8: Els angles a considerar són els més petits (α i β), tant si el vèrtex és de grau 2 com si és de grau > 2 . Una fita superior trivial de l'angle obtingut és $(2\pi/\text{grau_vèrtex})$ rad.

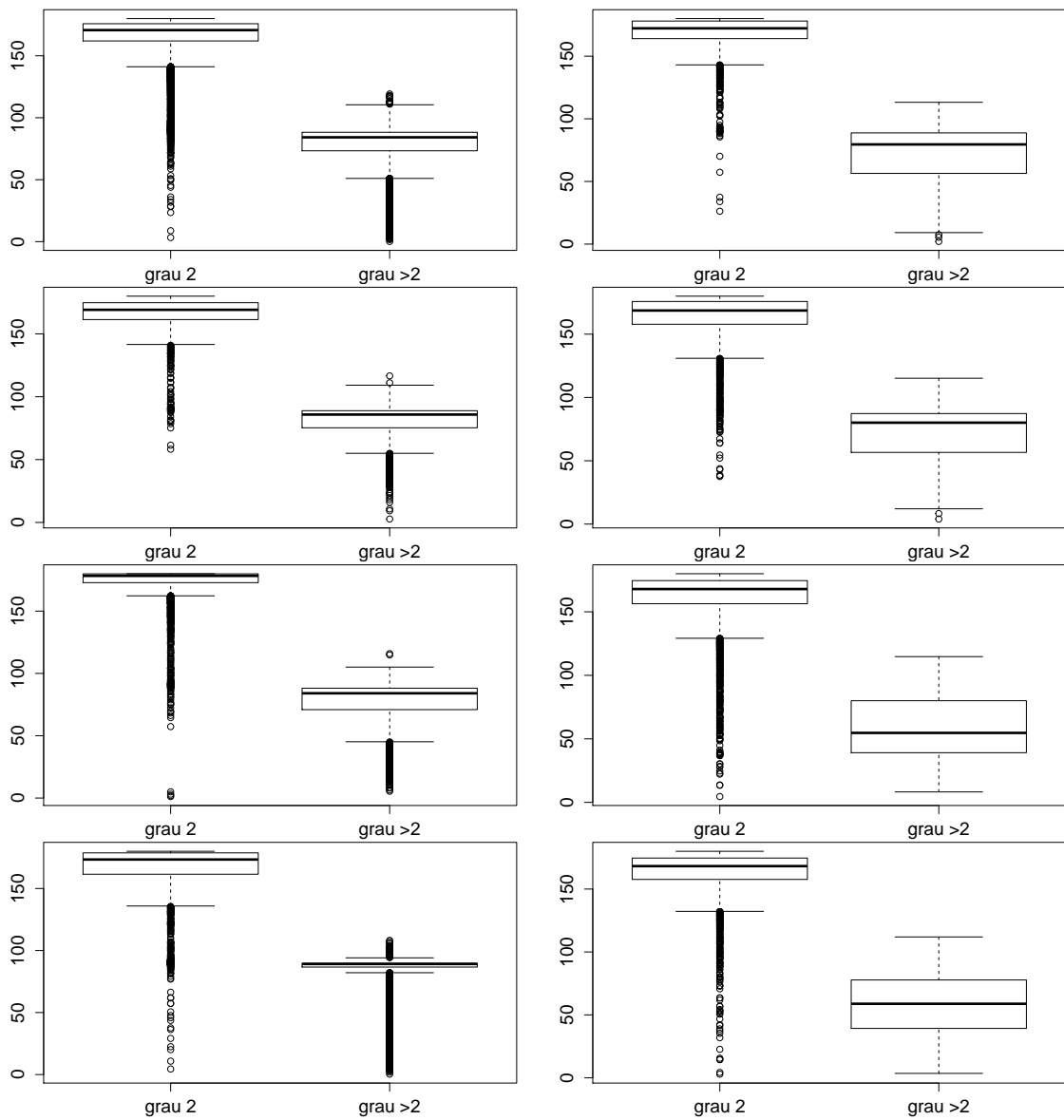
El llinard angular descrit en Karagiorgou i Pfofer [31] d'alguna manera vol representar una barrera horitzontal en les gràfiques de la Figura 4.9, de tal manera que s'ignora tot allò que està per sobre (vegeu la Figura 4.10). Un llinard més baix considerarà angles més oberts com a girs i faria que hi hagués més falsos positius (alçaria la barrera en les gràfiques) mentre que un llinard massa alt no consideraria alguna intersecció tot generant un fals negatiu (baixaria la barrera en les gràfiques).

Convé destacar que ni en la ciutat ni en el senderisme els vèrtexs que són intersecció es poden separar angularment dels vèrtexs intermedis i, per consegüent, és impossible

evitar els falsos positius i els falsos negatius a la vegada amb només el criteri angular. Aquesta consideració resulta evident en tenir en compte que hi poden haver corbes sense necessàriament haver-hi una intersecció. Tanmateix, tal com es veu en la [Figura 4.9](#), les interseccions són completament separables angularment des del 25è percentil dels vèrtexs intermedis. Més concretament, fixant-nos amb la [Taula 4.7](#), veiem que es pot separar els dos conjunts angularment tenint com a falsos positius menys del 10% dels vèrtexs intermedis.

Una observació interessant de les gràfiques de la [Figura 4.9](#) i la [Taula 4.7](#) és que el criteri angular per identificar interseccions sembla igual de vàlid en el senderisme com ho és en les ciutats. Tot i així, aquesta anàlisi s'ha fet sobre el *ground-truth*, mentre que l'algorisme aplica el llindar a les trajectòries. Tenint present aquest fet, convé destacar que en les trajectòries l'error del GPS té una incidència molt major en els angles del senderisme que en els de la ciutat, ja que la distància recorreguda entre dos mostres és molt més gran en aquest darrer (tal i com es mostra a la [Figura 4.7](#)), fent que el criteri angular no funcioni de la mateixa manera en ciutat que en senderisme.

Finalment, cal recordar que l'angle calculat per les interseccions és el més petit i, per tant, és el més estricte. Dit d'una altra manera, pot ser que hi hagi trajectòries que passin per la intersecció però no segueixin el gir més tancat i, conseqüentment, l'algorisme no identifiqui la intersecció. No obstant, això no representa un problema seriós quan els conjunts de dades són suficientment grans.



(a) Ciutats, de dalt a baix: *Atenes gran*, *Atenes petit*, *Berlín*, *Chicago*.

(b) Zones senderisme, de dalt a baix: *Delta del Llobregat*, *Aiguamolls Baix Empordà*, *Turó de l'home*, *Garraf*.

Figura 4.9: *Boxplots* dels diferents conjunts de dades. Aquestes gràfiques representen els angles mínims dels diferents vèrtexs agrupats segons si són vèrtexs intermedis (grau 2) o interseccions (grau > 2).

Lloc	Vèrtexs	Mitjana	StDev	Màx. (%)
ATENES GRAN				
<i>grau</i> = 2	18 866	165.80	16.64	119.21 (2.90 %)
<i>grau</i> > 2	12 160	77.54	17.06	
ATENES PETIT				
<i>grau</i> = 2	1458	164.33	17.98	116.59 (3.77 %)
<i>grau</i> > 2	1137	79.12	16.07	
BERLÍN				
<i>grau</i> = 2	4175	171.97	17.36	115.87 (2.92 %)
<i>grau</i> > 2	1507	76.32	18.26	
CHICAGO				
<i>grau</i> = 2	5226	162.98	26.34	108.17 (9.49 %)
<i>grau</i> > 2	3716	80.79	19.94	
DELTA DEL LLOBREGAT				
<i>grau</i> = 2	2813	168.27	14.97	113.18 (1.64 %)
<i>grau</i> > 2	353	69.57	24.60	
BAIX EMPORDÀ				
<i>grau</i> = 2	2960	162.47	20.20	115.18 (4.32 %)
<i>grau</i> > 2	502	70.69	22.90	
TURÓ DE L'HOME				
<i>grau</i> = 2	6790	161.57	20.71	114.74 (3.70 %)
<i>grau</i> > 2	71	58.65	27.11	
GARRAF				
<i>grau</i> = 2	4058	162.70	18.87	111.95 (2.76 %)
<i>grau</i> > 2	176	57.23	24.66	

Taula 4.7: Taula resum de dades importants en les comparacions angulars (en graus) dels vèrtexs intermedis amb les interseccions. La columna *Màx. (%)* indica l'angle màxim de les interseccions i a quin percentil de vèrtexs intermedis correspon aquest valor, és a dir, és el valor per al qual el llindar angular no genera falsos negatius i genera el mínim nombre de falsos positius (especificat en forma de percentatge sobre el total de vèrtexs de grau 2).

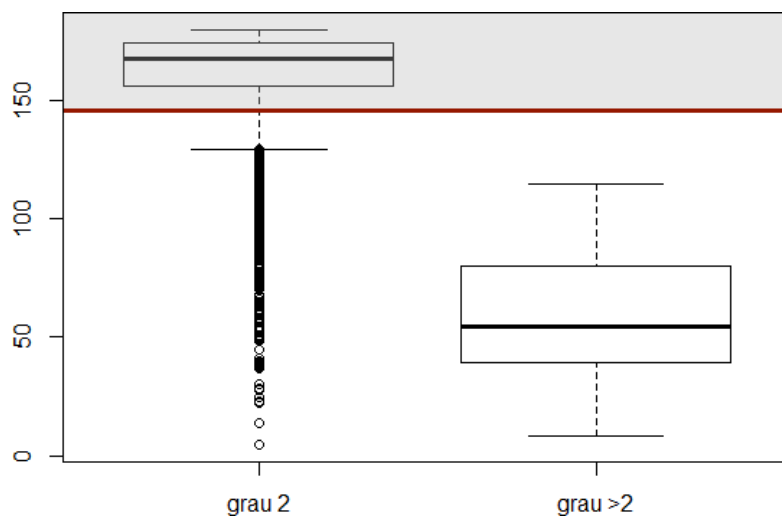


Figura 4.10: Un llindar angular d'uns 35° (representat per la línia de color vermell fosc) eliminaria tots els vèrtexs de l'àrea ombrejada. Si el llindar es posa massa avall hi haurà falsos negatius (interseccions ombrejades) i si, en canvi, es posa massa amunt hi haurà falsos positius (vèrtexs intermedis no ombrejats).

4.2.4 Mapes generats

Els mapes definitius generats pels algorismes a partir dels paràmetres triats (recordeu les [Taula 4.4](#) i [Taula 4.5](#)) es mostren a la [Figura 4.11](#) de forma global i juntament amb les trajectòries d'entrada. A la [Taula 4.8](#) s'hi pot trobar un resum de les seves complexitats.

Pel què fa referència a les complexitats dels resultats, cal remarcar que els mapes generats en senderisme són unes 10 vegades més complexos que a la ciutat en quant a nombre de vèrtexs per kilòmetres coberts. Una de les possibles causes d'aquest fet és que els camins en senderisme tenen, de per si, una geometria més complexa que la majoria de carrers i carreteres de ciutat de forma gairebé rectilínia.

En general, l'algorisme de Karagiorgou i Pfofer [\[31\]](#) genera més vèrtexs que el d'Ahmed i Wenk [\[4\]](#) tant en ciutat com en senderisme. Nogensmenys, en el senderisme, l'algorisme de Karagiorgou i Pfofer [\[31\]](#) té més tendència a crear artefactes (explicat en detall a [Secció 4.3](#)) que el d'Ahmed i Wenk [\[4\]](#) i, per tant, agafant el criteri de la freqüència i gravetat dels artefactes fabricats, els mapes Ahmed i Wenk [\[4\]](#) són globalment més bons, en els nostres conjunts de dades de senderisme, que els de Karagiorgou i Pfofer [\[31\]](#). Amb tot, depenent de les característiques que hom consideri com a desitjables, els mapes generats per Karagiorgou i Pfofer [\[31\]](#) poden ser més adequats ja que, per exemple, l'algorisme d'Ahmed i Wenk [\[4\]](#) és més sensible a *outliers*, la qual cosa fa que en els seus mapes puguin aparèixer camins sencers completament erronis.

Finalment, en comparar els mapes generats entre les zones de senderisme semblants a ciutat (*Delta del Llobregat* i *Aiguamolls Baix Empordà*) i les zones muntanyoses (*Turó de l'home* i *Garraf*), es fa evident un detriment de la qualitat d'aquests en els massissos. Concretament, les trajectòries en les zones muntanyoses contenen més soroll i, a més, els camins en aquestes zones són propers els uns als altres i presenten nombroses corbes. Aquests fets combinats fan que el nombre d'artefactes fabricats en zones de muntanya sigui més elevat que en les zones de senderisme pla. A la [Figura 4.12](#) s'exemplifiquen els mapes generats en una zona sorollosa del *Turó de l'home* en front d'una zona de camps pròpia dels *Aiguamolls Baix Empordà*.

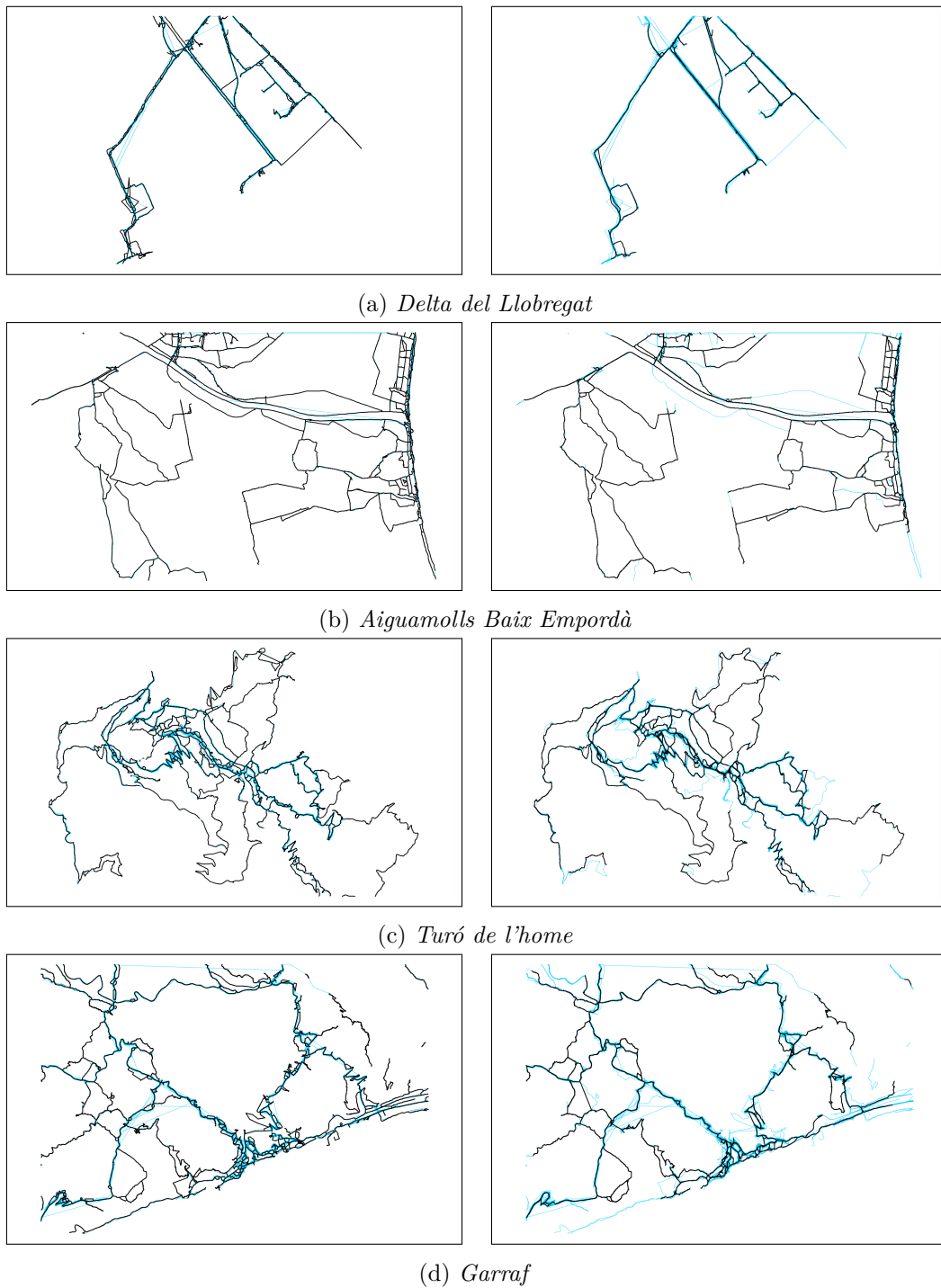
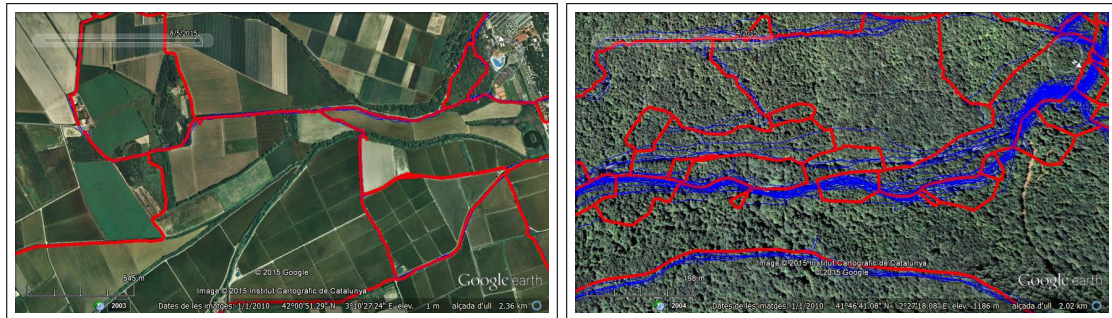


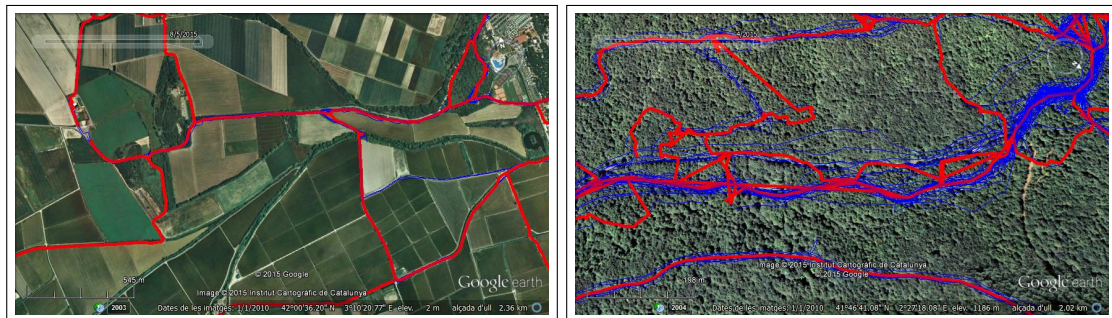
Figura 4.11: La primera i segona columnes corresponen als algorismes d'Ahmed i Wenk [4] i de Karagiorgou i Pfoser [31], respectivament. En negre surten els mapes generats i en blau les trajectòries.

Mapa generat	# Vèrt.	# Arest.	Long. [km]
ATENES GRAN			
Ahmed i Wenk	7067	7960	1358
Karagiorgou i Pfofer	6584	5280	252
ATENES PETIT			
Ahmed i Wenk	344	378	35
Karagiorgou i Pfofer	660	637	35
BERLÍN			
Ahmed i Wenk	1322	1567	164
Karagiorgou i Pfofer	2542	2262	161
CHICAGO			
Ahmed i Wenk	1195	1286	34
Karagiorgou i Pfofer	596	558	26
DELTA DEL LLOBREGAT			
Ahmed i Wenk	2546	2811	30
Karagiorgou i Pfofer	6619	6671	20
AIGUAMOLLS BAIX EMPORDÀ			
Ahmed i Wenk	10 822	11 127	129
Karagiorgou i Pfofer	23 020	23 134	108
TURÓ DE L'HOME			
Ahmed i Wenk	9813	10 166	104
Karagiorgou i Pfofer	28 547	28 714	111
GARRAF			
Ahmed i Wenk	9379	9727	106
Karagiorgou i Pfofer	35 231	35 331	77

Taula 4.8: Complexitats dels mapes generats en els conjunts de dades tant de ciutat (fent servir les dades de Ahmed et al. [5]) com de senderisme.



(a) Ahmed i Wenk [4]



(b) Karagiorgou i Pfoer [31]

Figura 4.12: Il·lustració de com el soroll en les trajectòries, propi de zones de muntanya, afecta la qualitat del mapa generat. En vermell els mapes generats i en blau les trajectòries. La primera columna correspon als *Aiguamolls Baix Empordà* i la segona al *Turó de l'home*. Imatges extretes de Google Earth.

4.3 Artefactes generats

En aquesta secció s'analitzen qualitativament els artefactes que creen els algorismes quan s'executen amb dades de senderisme (*i. e.* construccions que no segueixen la geometria subjacent de les trajectòries d'entrada). Aquests *artefactes* els podem classificar segons el seu origen en si són *dependents dels paràmetres* o bé *inherents a l'algorisme*.

Els artefactes *dependents dels paràmetres* són aquells que poden o no donar-se en una regió concreta segons els valors dels paràmetres. En canvi, els artefactes *inherents a l'algorisme* són conseqüència del propi algorisme i, malgrat no tenen perquè donar-se en tots els conjunts de dades, quan succeeixen no depenen del valor dels paràmetres. Cal dir que, sobre els artefactes *dependents dels paràmetres*, malgrat es pugui trobar valors dels paràmetres perquè localment no succeeixin en una regió determinada, això no significa que existeixi una parametrització que els elimini de forma global en tot el mapa generat.

4.3.1 Ahmed i Wenk

Artefactes dependents dels paràmetres

Un dels artefactes més freqüents d'Ahmed i Wenk [4] és quan dos camins paral·lels es troben a una distància llindar la qual depèn d' ε . En aquesta situació, l'algorisme tendirà a unir i separar aquests dos camins de forma freqüent, creant així petits «quadratets» entre ells (vegeu la [Figura 4.13](#)).

En un altre extrem, si el paràmetre ε és massa alt, una intersecció en forma de H pot resultar en què, si el camí del mig té longitud menor a 2ε , aquest no aparegui en el mapa generat. El motiu és que, a causa de la naturalesa incremental de l'algorisme, pot ser que s'identifiquin primer els dos camins paral·lels abans que qualsevol trajectòria que passi pel camí del mig. Si aquest és el cas, totes les trajectòries del camí del mig queden col·lapsades en els altres dos camins, malgrat pugui ser que els intersequin de forma gairebé ortogonal. A la [Figura 4.14](#) s'explica gràficament aquest fet i a la [Figura 4.15](#) se'n mostra un exemple.

Finalment, un altre artefacte que depèn del paràmetre ε és la unió de camins a interseccions. Aquest algorisme presenta un efecte de «drecera» quan dos camins intersequen amb angles tancats. En aquest cas, incrementar ε fa que la drecera es produeixi

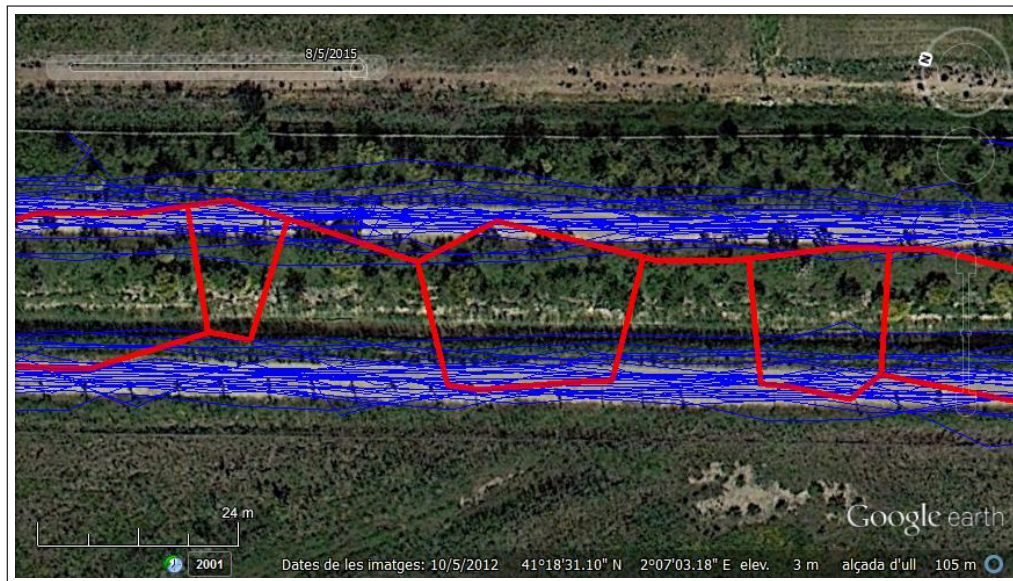


Figura 4.13: Efecte causat quan dos camins estan a una distància llindar d' ϵ . En blau hi ha les trajectòries i en vermell el mapa generat. Imatge extreta de Google Earth en el *Delta del Llobregat*.

abans, agreujant-ne el problema. A la Figura 4.16 hi ha un exemple d'aquest fenomen.

Artefactes inherents a l'algorisme

Un dels inconvenients principals de l'algorisme d'Ahmed i Wenk [4] és la seva sensibilitat a *outliers*. L'algorisme manté internament un graf amb tota la geometria analitzada fins el moment i hi insereix noves trajectòries incrementalment. A causa de la manca de cap poda de trajectòries inserides així com la falta de cap control sobre la densitat de trajectòries que mostregen un mateix camí, s'esdevenen dues conseqüències:

1. Tots els *outliers* que s'allunyin més que ϵ de totes les altres trajectòries, per qualsevol ϵ raonable, acabaran formant part del mapa generat. Aquest efecte està exemplificat a la Figura 4.17.
2. Hi haurà trossos de camins molt densos pels quals no hi haurà camí en el mapa generat, ja que una trajectòria errònia tindrà distància menor a ϵ als trossos en

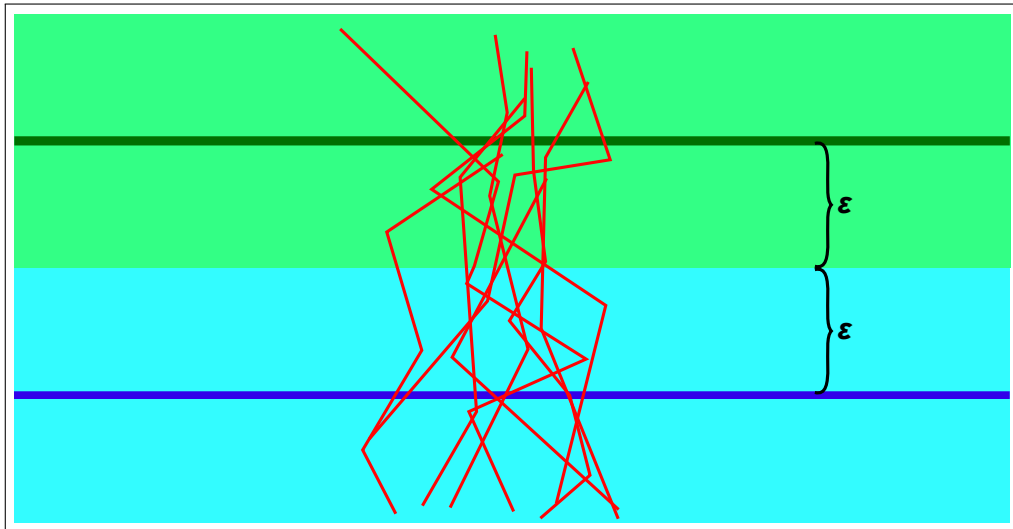


Figura 4.14: Esquema mostrant l'efecte d'una ε (localment) massa gran. Suposant que les rectes verda i blava formen part del mapa incremental, la inserció de les trajectòries vermelles no tindrà cap efecte ja que aquestes no surten de l'àrea d'«influència» d'aquestes dues rectes, tot i ser ortogonals a aquestes.

qüestió i, per la naturalesa incremental de l'algorisme, anul·larà totes les altres trajectòries. Aquest efecte està il·lustrat a [Figura 4.18](#).

Per acabar, l'últim artefacte identificat d'aquest algorisme succeeix quan hi ha trajectòries que passen múltiples cops pel mateix camí. Un exemple clar són les trajectòries d'«anada i tornada», molt freqüents en el senderisme, per exemple, en una equivocació del camí o en la visita d'un lloc d'interès paisatgístic. Quan l'algorisme insereix aquest tipus de trajectòries no comprova que en la pròpia trajectòria hi hagi regions properes de menys d' ε i, per tant, s'afegeix per duplicat la geometria pertanyent a un mateix camí (vegeu la [Figura 4.19](#)).

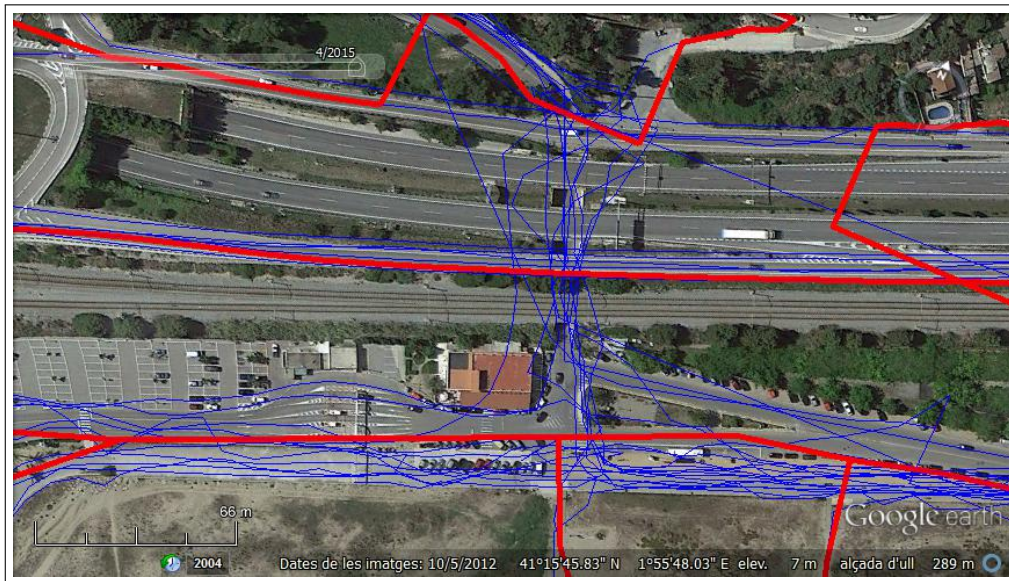


Figura 4.15: Intersecció en forma de H sense que es generi el camí del mig a causa d'una ε (localment) massa gran. En blau hi ha les trajectòries i en vermell el mapa generat. Imatge extreta de Google Earth en el *Garraf*.

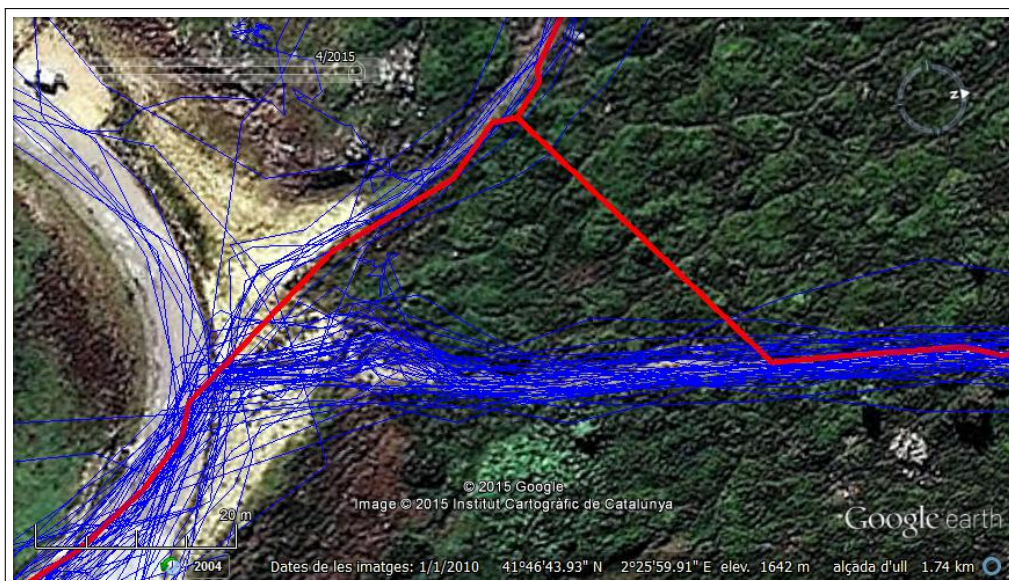


Figura 4.16: Imatge il·lustrant el fenomen de «drecera» causat per una ε alta. En blau hi ha les trajectòries i en vermell el mapa generat. Imatge extreta de Google Earth en el *Turó de l'home*.

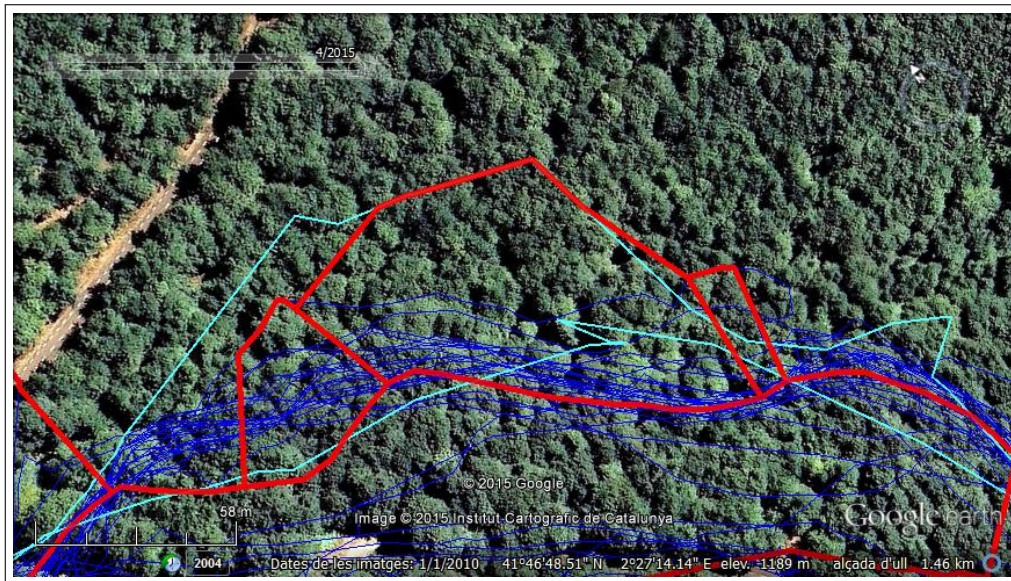


Figura 4.17: Exemple de zones de mapa creat a causa d'un *outlier*. En blau hi ha les trajectòries, en cian l'*outlier* i en vermell el mapa generat. Imatge extreta de Google Earth en *Turó de l'home*.

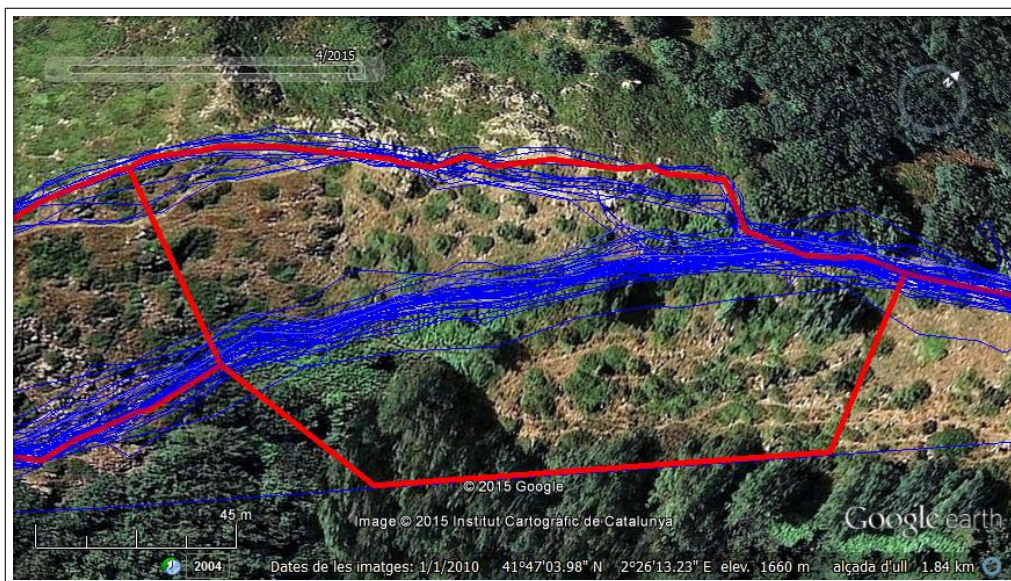


Figura 4.18: Exemple de trossos densament mostrejats per les trajectòries però que no han generat geometria a causa d'un *outlier* que ha sigut inserit anteriorment. Malgrat la distància aparent entre els camins, cal recordar que en la imatge l'alçada juga un paper important que l'algorisme ignora. En blau hi ha les trajectòries i en vermell el mapa generat. Imatge extreta de Google Earth en el *Turó de l'home*.



Figura 4.19: Exemple de trajectòria d'«anada i tornada». Aquest tipus de trajectòries insereixen per duplicat geometria corresponent al mateix camí. En vermell hi ha el mapa generat que se sobreposa a la única trajectòria d'aquesta zona. Imatge extreta de Google Earth en els *Aiguamolls Baix Empordà*.

4.3.2 Karagiorgou i Pfofer

Artefactes dependents dels paràmetres

La detecció d'interseccions a partir de girs és el pilar fonamental d'aquest algorisme. Els efectes de no poder detectar bé les interseccions a causa d'una mala parametrització (local) són transcendents en la qualitat dels resultats i es poden classificar en dos tipus:

- La no detecció de girs (i per tant d'interseccions) fa que camins amb densitat de trajectòries no siguin representats en el mapa generat (vegeu la [Figura 4.20](#)).
- Un excés de girs detectats provoca la creació d'interseccions no reals en la zona. A la [Figura 4.21](#) hi ha un exemple de com aquestes interseccions falses fan que el mapa generat no tingui només la forma de zig-zag del camí sinó que també apareguin dreceres a través.



Figura 4.20: Imatge sobre com la no detecció de girs causa la pèrdua de camins en el mapa generat. En vermell hi ha el mapa generat, en blau les trajectòries, les icones vermelles són els girs detectats i les icones grogues les interseccions que aquests girs generen. Imatge extreta de Google Earth en els *Aiguamolls Baix Empordà*.

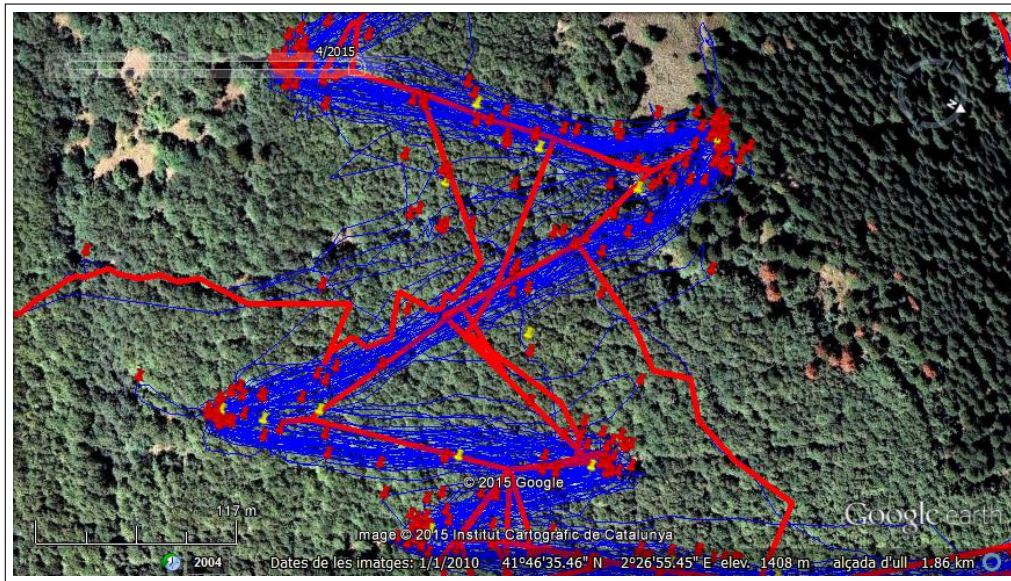


Figura 4.21: Il·lustració de l'efecte de generar interseccions falses en una zona en forma de zig-zag. Les interseccions falses són connectades entre elles creant camins artificials a través del zig-zag. En vermell hi ha el mapa generat, en blau les trajectòries, les icones vermelles són els girs detectats i les icones grogues les interseccions que aquests girs generen. Imatge extreta de Google Earth en el *Turó de l'home*.

Artefactes inherents a l'algorisme

L'algorisme de Karagiorgou i Pfoser [31] a petita escala genera un elevat nombre de geometria artificial. Aquests artefactes són fabricats en la fase d'agrupar les trajectòries per formar els enllaços entre interseccions. La causa de fons de la seva existència no ha pogut ser verificada ja que a l'article original no es mencionen i no s'ha pogut fer un anàlisi profund del codi per la manca de llegibilitat i l'extensió d'aquest. No obstant, s'ha pogut observar com zones amb un elevat grau de soroll són més propenses a contenir aquests artefactes i, per tant, els mapes generats en les zones de senderisme (sobretot zones com el *Turó de l'home* i el *Garraf*) en pateixen notòriament més que zones de ciutat.

Els tres artefactes més paradigmàtics causats pel sistema d'agrupament de trajectòries són:

1. Influència de trajectòries properes però de camins diferents, fent que els dos camins del mapa generat tinguin tendència a, erròniament, apropar-se un amb l'altre. Un exemple il·lustratiu es troba a la [Figura 4.22](#).
2. Geometria estranya com bucles o camins no monòtons. A la [Figura 4.23](#) s'hi pot trobar un exemple d'un bucle al conjunt de proves original de *Atenes*, així com un exemple en el *Garraf*.
3. En zones d'elevat soroll es generen molts artefactes creant una geometria completament errònia (vegeu la [Figura 4.24](#)).



Figura 4.22: Exemple de com les trajectòries de dos camins propers s'influencien unes amb les altres, fent que els camins del mapa generat no siguin els correctes. En vermell hi ha el mapa generat, en blau les trajectòries, les icones vermelles són els girs detectats i les icones grogues les interseccions que aquests girs generen. Imatge extreta de Google Earth en el *Delta del Llobregat*.

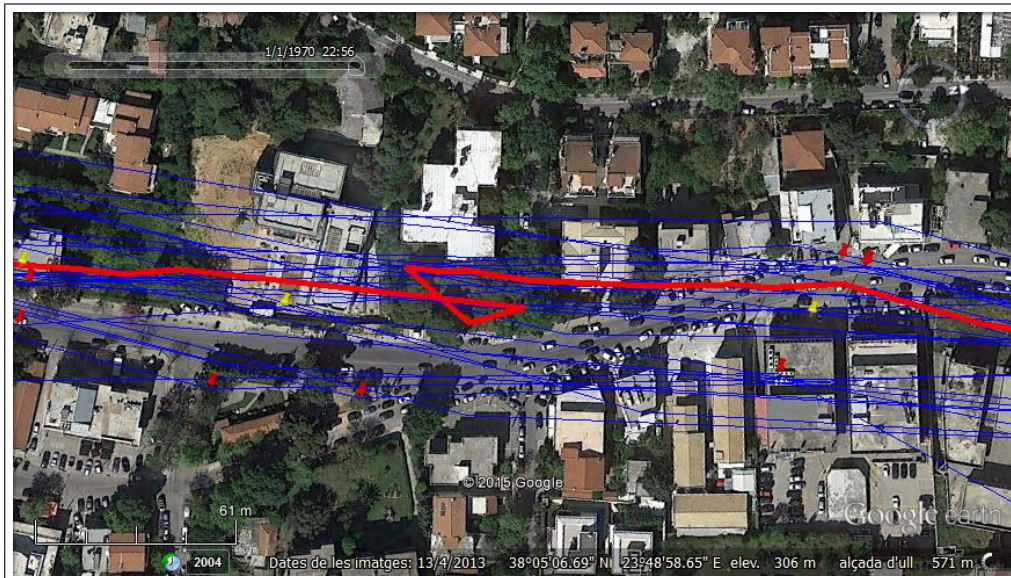
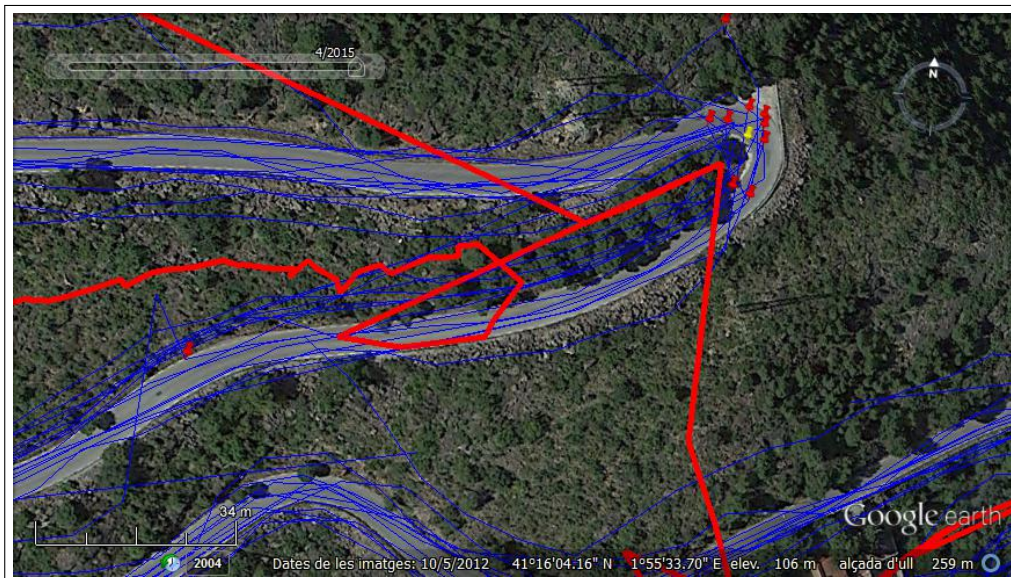
(a) *Atenes petit*(b) *Garraf*

Figura 4.23: Il·lustració de bucles a l'hora de fusionar les diferents trajectòries. En vermell hi ha el mapa generat, en blau les trajectòries, les icones vermelles són els girs detectats i les icones grogues les interseccions que aquests girs generen. Imatges extretes de Google Earth.

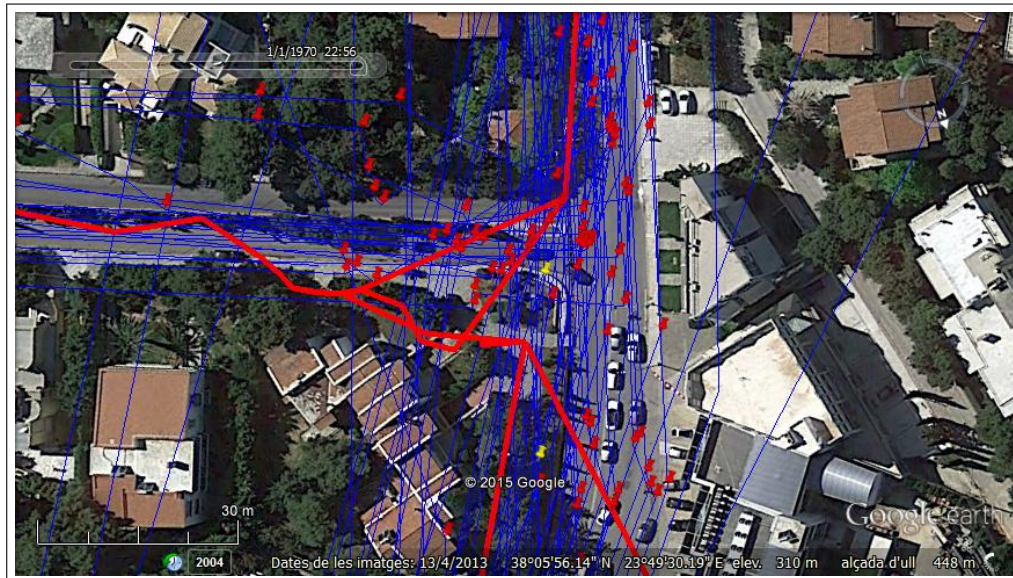
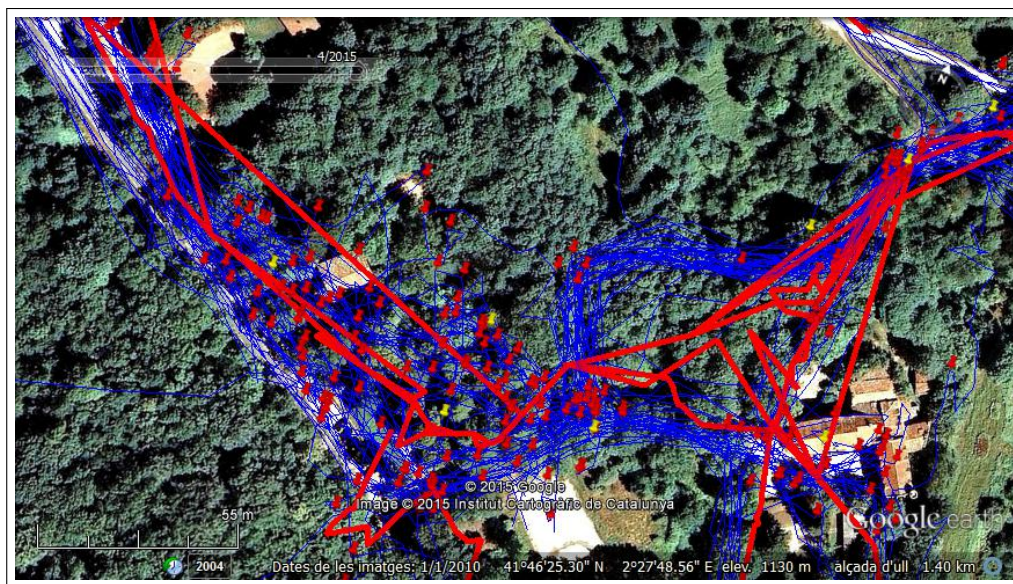
(a) *Atenes petit*(b) *Turó de l'home*

Figura 4.24: En vermell hi ha el mapa generat, en blau les trajectòries, les icones vermelles són els girs detectats i les icones grogues les interseccions que aquests girs generen. Imatges extretes de Google Earth.

4.4 Punts forts i mancances dels algorismes en el senderisme

A partir dels resultats de l'anàlisi qualitatiu dels mapes generats, en aquesta secció es fa una valoració de les idees d'ambdós algorismes classificant-les segons la seva adequació en el senderisme.

La idea fonamental d'Ahmed i Wenk [4], compartida per tots els algorismes de generació de mapa, és la idea de «proximitat». Mentre que l'algorisme d'Ahmed i Wenk [4] es preocupa per la proximitat entre corbes utilitzant la distància Fréchet, l'algorisme de Karagiorgou i Pfoser [31] se centra en la distància Euclidiana entre punts (girs i interseccions). Ambdós algorismes, però, treballen en el pla, ja que aquest és suficient per la generació de mapes en ciutats i, fins i tot, en zones de senderisme pla com és el *Delta del Llobregat* i els *Aiguamolls Baix Empordà*. Tanmateix, restringir la geometria a dues dimensions suposa, en massissos com el *Turó de l'home* i el *Garraf*, una disminució no menyspreable de la «proximitat» dels elements (vegeu la [Figura 4.25](#)).

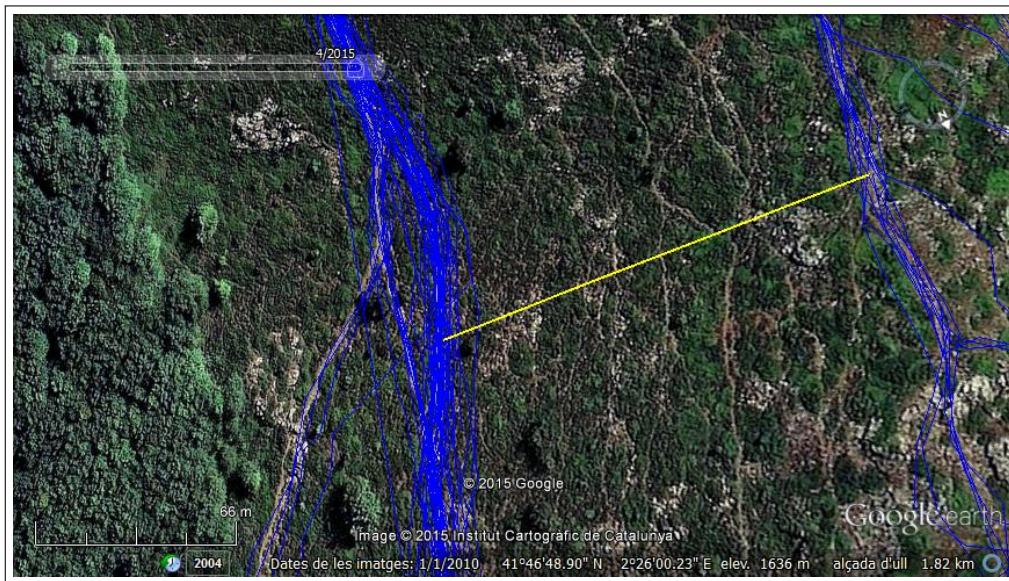


Figura 4.25: La distància de la recta groga en el pla és 86.71 m, mentre que considerant l'alçada aquesta esdevé 95.58 m, un 10.23% més. En blau es troben les trajectòries que indiquen la presència de dos camins, un que passa pel vessant de la muntanya i l'altre per la crinera. Imatge extreta de Google Earth en el *Turó de l'home*.

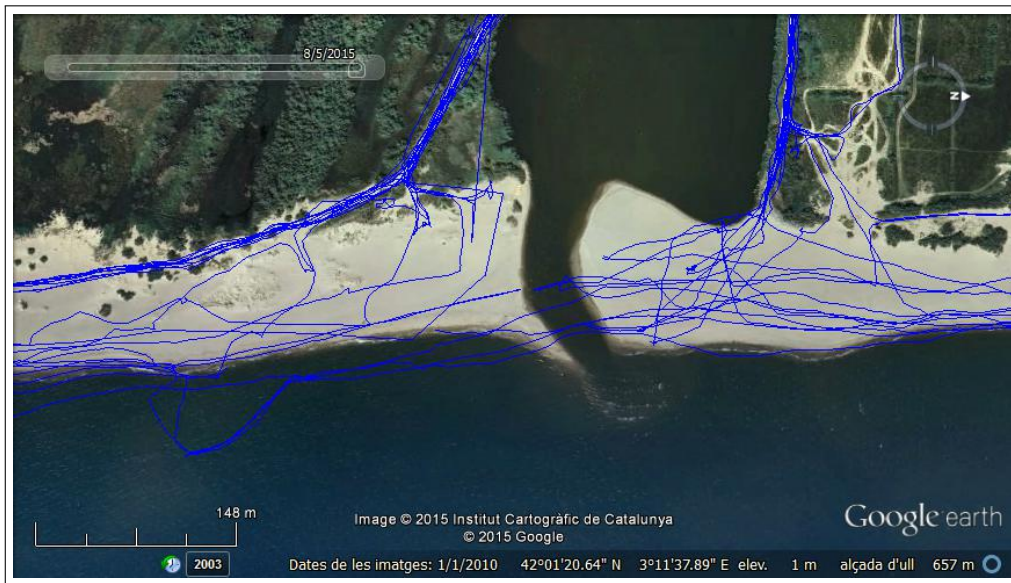
Una característica important en el senderisme és la heterogeneïtat dels diferents camins que s'utilitzen, és a dir, en una mateixa zona hi poden haver des de un camí ample o

esplanada sense camí marcat, a camins estrets i accidentats pel vessant d'una muntanya. Aquesta heterogeneïtat, sobretot si es consideren zones mixtes de pla i muntanya, fa que inevitablement la parametrització dels algorismes no sigui l'adequada simultàniament per a tota la zona. Conseqüentment, en el senderisme és important que l'algorisme sigui flexible i, o bé identifiqui els diferents tipus de camins, o bé admeti diverses parametritzacions adequades pels diferents trossos de mapa. A la [Figura 4.26](#) es mostra com en un mateix conjunt de dades de senderisme una sola parametrització és insuficient per funcionar correctament en la diversitat de situacions que conté.

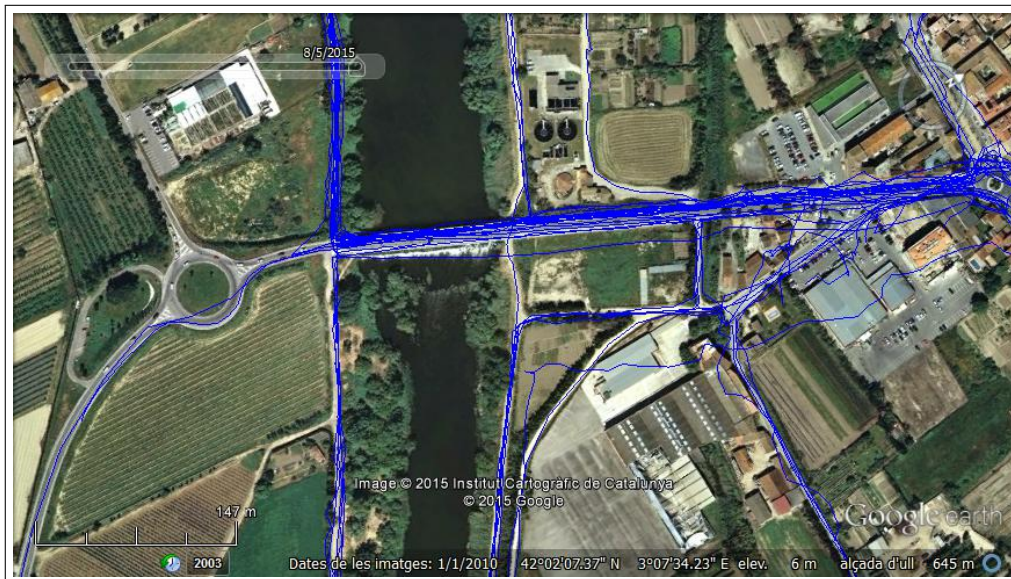
Sobre la idea de deteccions d'interseccions de Karagiorgou i Pfoer [31], basada en detectar primer els girs i agrupar-los, presenta dos inconvenients majors. Primerament, la distinció de girs a partir de velocitat no aplica a senderisme. En segon lloc, la distinció angular plantejada es veu greument afectada pel soroll elevat en els angles causat per la poca distància en els punts mostrejats (recordeu la [Figura 4.7](#)). Una forma d'alleujar el soroll és no considerar el gir angular que formen els nodes de cada trajectòria, sinó considerar el gir angular en termes de distància, a fi que les variables de temps de mostreig i velocitat, heterogènies, ja no introdueixin soroll a la detecció dels girs (vegeu la [Figura 4.27](#)).

Un altre punt destacable és la filtració d'*outliers*, preprocessat de gran importància en algorismes sensibles a ells (un cas paradigmàtic és Ahmed i Wenk [4]). En l'article original de Karagiorgou i Pfoer [31] es menciona l'ús del temps de mostreig com a filtre mentre que Ahmed i Wenk [4] no fan cap menció a mètodes de filtratge. En els codis, tanmateix, Ahmed i Wenk [4] usen el temps de mostreig per filtrar i Karagiorgou i Pfoer [31] una distància màxima. Donada la poca fiabilitat de les marques de temps (n'hi ha que no són ni seqüencials en el temps), l'enfocament de filtrar salts espacials partint la trajectòria en dos sembla més apropiat.

Finalment, un altre preprocessat que esdevé important és la segmentació de trajectòries d'«anada i tornada» (freqüents en senderisme) per tal d'evitar que una trajectòria mostregi múltiples cops un mateix camí. Algorismes com Ahmed i Wenk [4] sense aquest preprocessat produeixen resultats erronis i, per l'altre banda, algorismes com Karagiorgou i Pfoer [31] que contenen models prefixats de girs (en el qual no s'inclouen anar marxa enrere) poden tenir dificultats en tractar aquests casos. A més, cal remarcar que ni Ahmed i Wenk [4] ni Karagiorgou i Pfoer [31] consideren maneres de filtrar-los a causa de, probablement, la seva rara ocurrència en ciutat.



(a) Zona de platja



(b) Zona entrant a ciutat

Figura 4.26: La imatge (a) conté una zona de platja ample per la qual les trajectòries es mouen sense camí prefixat. En canvi, a la imatge (b) es mostra una zona de camins marcats i propers uns als altres (ambdues amb la mateixa escala). La parametrització adequada diferirà en (a) que en (b) malgrat pertanyin al mateix conjunt de dades dels *Aiguamolls Baix Empordà*. Imatges extretes de Google Earth.

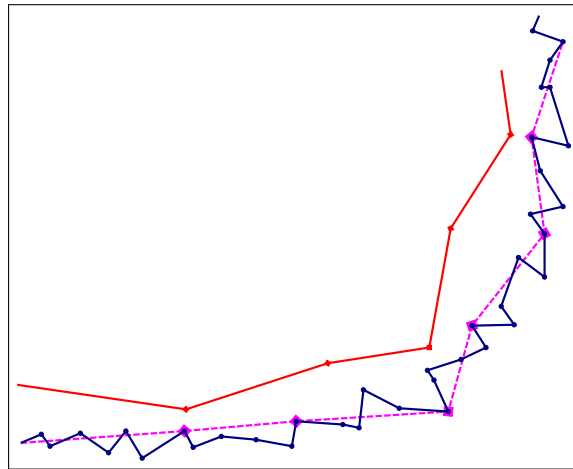


Figura 4.27: Esquema il·lustrant els beneficis de considerar diferències angulars cada certa distància i no a cada node. Les trajectòries vermella i blava tenen temps de mostreig molt diferents (assumint mateixa velocitat) fent que la blava contingui molt soroll angular. En canvi, considerant les desviacions angulars a distàncies similars, la trajectòria vermella i rosa contenen un soroll angular similar.

4.5 Mesures de qualitat

Més enllà de l'anàlisi qualitatiu per inspecció visual dels diferents mapes generats, també s'han trobat les distàncies Hausdorff dirigides [7], les distàncies basades en camins [3] i les distància de camins més curts [31], seguint l'exemple de Ahmed et al. [5].

Malgrat fora dels objectius d'aquest treball, s'han calculat els temps d'execució d'aquestes mesures de qualitat que són, en mitjana, 41.59 min, 1.04 h i 12 h en el càlcul de la distància Hausdorff dirigida, la distància basada en camins i la distància de camins curts, respectivament. Com es pot observar, intentar parametritzar els algorismes a fi de minimitzar aquestes mesures de qualitat és computacionalment taxatiu, fet que ha esdevingut un dels motius pels quals la parametrització s'ha realitzat via inspecció visual.

4.5.1 Distància Hausdorff dirigida i distància basada en camins

Per la distància Hausdorff dirigida, s'han computat tots els camins de longitud 1 (descartant vèrtexs de grau 2) del mapa generat i, per cada camí, s'ha calculat la distància Hausdorff dirigida entre la unió de les seves arestes originals i el *ground-truth*. Per la distància basada en camins, s'han computat els camins de longitud 3 del mapa generat i, per cada camí, s'ha calculat la distància Fréchet entre ell i la porció corresponent del *ground-truth* trobada per *map-matching* [8]. A partir d'ambdues distàncies per separat, s'han calculat el mínim, el màxim, la mediana, la mitjana, i els màxims traient el $d\%$ més alts (*outliers*), per a $d = 2, 5, 10$ i 15 . A la taula [Taula 4.9](#) es mostren els resultats obtinguts juntament amb els de Ahmed et al. [5].

En general, Karagiorgou i Pfoser [31] obté més bons resultats que Ahmed i Wenk [4] tant en la distància Hausdorff dirigida com en la distància basada en camins. Les tècniques d'agrupació d'arestes de Karagiorgou i Pfoser [31] semblen capaces de produir arestes de més qualitat i de reduir el nombre d'*outliers* amb comparació amb Ahmed i Wenk [4], on la qualitat de les arestes depèn de la qualitat de les trajectòries de forma individual i no hi ha gaires mesures de filtració d'*outliers*.

En la comparació entre la qualitat en zones de ciutat respecte zones de senderisme, només la zona del *Delta del Llobregat* és directament comparable als conjunts de dades de ciutat. Tant la distància Hausdorff dirigida com la distància basada en camins estan pensats per treballar sota la hipòtesi que el *ground-truth* cobreix tot el terreny, hipòtesi

Lloc	Distància Hausdorff dirigida (m)								Distància basada en camins (m)							
	mín.	màx.	med.	mitj.	2%	5%	10%	15%	mín.	màx.	med.	mitj.	2%	5%	10%	15%
ATENES GRAN																
Ahmed	1	269	30	33	84	67	56	50	7	849	70	85	250	164	132	114
Karagiorgou	1	200	10	13	46	35	26	22	2	175	25	32	109	80	63	53
ATENES PETIT																
Ahmed	1	82	25	26	82	54	46	40	9	224	45	52	101	101	81	72
Karagiorgou	2	84	14	17	54	40	33	30	7	229	32	38	113	68	59	57
BERLÍN																
Ahmed	1	219	30	33	95	70	60	53	9	540	66	74	207	147	120	107
Karagiorgou	1	232	14	18	59	42	34	30	4	306	28	37	120	85	65	52
CHICAGO																
Ahmed	1	81	14	19	72	59	43	35	7	201	35	42	127	100	85	76
Karagiorgou	1	48	7	8	41	23	15	13	3	89	15	23	72	72	65	51
DELTA																
Ahmed	1	2401	8	37	58	37	25	19	2	191	21	30	117	89	63	50
Karagiorgou	2	64	9	12	44	32	16	14	3	191	16	25	109	64	42	34
BAIX EMPORDÀ																
Ahmed	1	2401	24	58	189	166	143	119	3	1033	81	104	351	261	216	187
Karagiorgou	1	209	20	51	178	160	144	128	2	1034	80	99	305	233	207	169
TURÓ DE L'HOME																
Ahmed	1	2401	29	81	371	229	158	110	3	1408	68	111	602	327	234	181
Karagiorgou	2	1177	20	55	247	213	151	119	3	1406	67	112	537	310	237	214
GARRAF																
Ahmed	1	2401	17	53	293	141	79	59	3	594	53	82	405	279	185	127
Karagiorgou	1	527	14	28	189	131	53	36	3	868	42	77	430	308	188	113

Taula 4.9: Resultats de la mesura quantitativa de qualitat de la distància basada en camins i la distància Hausdorff dirigida. S'inclouen les dades de Ahmed et al. [5] i les dels nostres conjunts de dades.

certa només en el *Delta del Llobregat*. Tal i com ja s'ha mostrat a la [Figura 3.8](#), en els *Aiguamolls Baix Empordà* al *ground-truth* li mancava la zona corresponent a la platja, mentre que en el *Turó de l'home* i el *Garraf* la zona coberta és més escassa i s'hi pot observar la mancança de múltiples camins. Amb tot, l'efecte en les mesures d'un *ground-truth* no complet es pot quantificar a partir dels increments de les distàncies entre el *Delta del Llobregat* i els *Aiguamolls Baix Empordà*, on les trajectòries es comporten de forma similar al ser aquestes zones de senderisme pla. La mitjana d'aquests increments en ambdues mesures i algorismes són del 230.75% en la mitjana i del 251.75% en la mediana.

Consegüentment, de les mesures de la distància Hausdorff dirigida i la distància basada en camins, només en podem extreure que els mapes generats en la zona de senderisme del *Delta del Llobregat* tenen millors resultats que en la pròpia ciutat però, no obstant,

aquestes mesures de qualitat no són aplicables en general al senderisme a causa de la manca de *ground-truth* amb bona cobertura del terreny.

4.5.2 Distància de camins més curts

Per la distància de camins més curts s'han computat 2000¹ camins més curts entre parelles de punts escollits en el mapa generat. Els dos punts de cada parella són triats, un d'ells, aleatòriament i l'altre s'agafa com el més llunyà en una finestra de 2 km d'aquest, tal i com es fa en el codi original. Per cada parella de punts, també es calcula el camí més curt corresponent en el *ground-truth*, i ambdós camins, el del mapa generat i el del *ground-truth*, es comparen mitjançant la distància Fréchet discreta i la distància mitjana vertical. Cal dir que part dels camins curts no s'han pogut calcular en el *ground-truth* per raó de problemes de connectivitat.

El percentatge de camins trobats juntament amb el mínim, màxim, mitjana i desviació estàndard de les distàncies Fréchet discreta, mitjana vertical i total del camí més curt calculat en el *ground-truth*; estan recollits a la [Taula 4.10](#), on també s'inclouen els resultats de Ahmed et al. [5].

En els conjunts de dades de senderisme, el percentatge de camins trobats decau considerablement i, per tant, també la rellevància estadística de les dades, *i. e.* es sobremostregen certs camins en front d'altres. Una de les causes d'aquest fet és l'agressivitat de la poda del *ground-truth*. El sistema d'eliminar arestes a partir de la presència del mapa generat prop de l'aresta pot produir falsos positius, sobretot quan les arestes són de mida reduïda, tal i com passa en el senderisme on els camins tenen geometria complexa (a la [Figura 4.28](#) se'n mostra un exemple). Amb tot, aquesta mesura és més sensible a les imperfeccions del *ground-truth* que la distància Hausdorff i la distància basada en camins, ja que petites variacions en el mapa poden generar camins curts completament diferents. Per consegüent, no es poden extreure conclusions d'aquestes estadístiques a causa de la mala cobertura del *ground-truth* i la poca adequació de la poda en mapes d'arestes curtes.

¹Ahmed et al. [5] n'han utilitzat 500.

Lloc	Trobat (%)	D. Fréchet discr. (m)				Dist. mitj. vert. (m)				Dist. camí curt (km)			
		mín.	màx.	mitj.	SD	mín.	màx.	mitj.	SD	mín.	màx.	mitj.	SD
ATENES GRAN													
Ahmed	92.6	23	445	137	103	12	230	106	62	1.12	11.84	6.93	2.92
Karagiorgou	94.2	19	432	125	96	9	225	98	58	1.01	11.62	6.84	2.86
ATENES PETIT													
Ahmed	97.6	13	234	96	62	6	91	38	24	1.28	5.72	3.11	1.84
Karagiorgou	96.8	7	212	81	48	3	81	27	20	0.78	5.21	2.95	1.39
BERLÍN													
Ahmed	93.2	21	469	191	123	12	231	121	63	1.56	5.88	3.49	1.96
Karagiorgou	93.8	18	428	183	112	8	209	106	58	1.32	5.67	3.27	1.84
CHICAGO													
Ahmed	99.8	13	208	97	56	6	92	43	19	1.21	6.95	4.45	2.04
Karagiorgou	99.2	4	103	41	28	2	50	21	14	0.90	6.05	3.82	1.59
DELTA													
Ahmed	8.9	201	612	264	89	18	180	43	39	2.56	6.64	4.81	1.00
Karagiorgou	2.1	230	402	291	65	30	110	64	28	2.31	7.10	4.80	1.36
BAIX EMPORDÀ													
Ahmed	13.0	94	2576	1193	832	3	1207	428	359	2.24	17.40	5.23	2.24
Karagiorgou	19.6	94	3538	766	992	3	1467	236	434	2.13	16.28	5.75	2.82
TURÓ DE L'HOME													
Ahmed	44.3	48	2074	704	500	7	825	239	200	2.95	15.14	5.64	2.14
Karagiorgou	58.1	48	1322	563	232	4	536	157	95	2.86	11.24	4.82	1.44
GARRAF													
Ahmed	46.0	123	3267	1324	973	16	2056	690	589	2.77	10.71	5.95	1.67
Karagiorgou	21.1	53	2786	252	320	3	1304	52	141	2.62	7.47	4.32	0.79

Taula 4.10: Resultats de la mesura quantitativa de qualitat de la distància de camins curts. S'inclouen les dades de Ahmed et al. [5] i les dels nostres conjunts de dades.

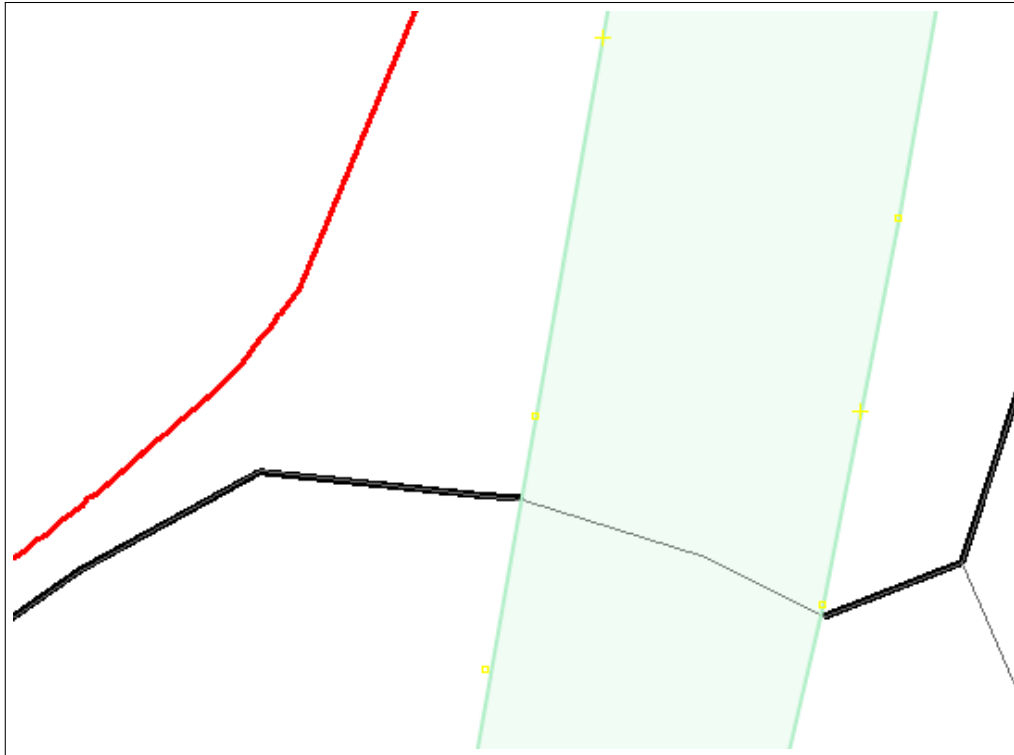


Figura 4.28: Il·lustració de com arestes de longitud petita poden quedar podades malgrat ser pròximes al mapa generat. En negre el *ground-truth* podat, en gris el *ground-truth* original, en vermell el mapa generat i en verd la geometria augmentada de l'aresta que manca ja que el mapa generat no hi interseca. Imatge treta del *Delta del Llobregat*.

5 Gestió del projecte

5.1 Planificació temporal

Duració projecte Aquest projecte està emmarcat dins d'un període molt concret de dates d'inici i de final, tal com es mostra a la [Taula 5.1](#), formant un període de treball de 133 dies, incloent-hi dies no laborables. Assumint una mitjana de treball diari de 3 h i mitja, equival a un període de treball de 464.5 hores.

Data Inici	Data Final
9 de febrer de 2015	22 de juny de 2015

Taula 5.1: Dates d'inici i final del projecte.

Diagrama de Gantt Dins d'aquest període de treball, s'identifiquen quatre grans blocs de tasques: *Gestió del projecte (GEP)*, *Establiment del framework*, *Experiments* i *Redacció memòria*. Per tenir una visió global i detallada de totes les tasques de cada un dels blocs s'ha fet ús d'un diagrama de Gantt dividit en part textual i part visual:

- Part textual: Hi trobem en forma de taula la data d'inici i fi de les fases, tasques, subtasques i *milestones*. Per cada subtasca s'indica la persona responsable de l'execució.
- Part visual: Hi trobem les tasques ordenades pel temps amb les seves dependències. S'ha pintat de color el risc que té cada una d'elles seguint el següent esquema de colors: *blau*, sense risc; *verd*, risc baix; *groc*, risc mitjà; *vermell*, risc elevat.

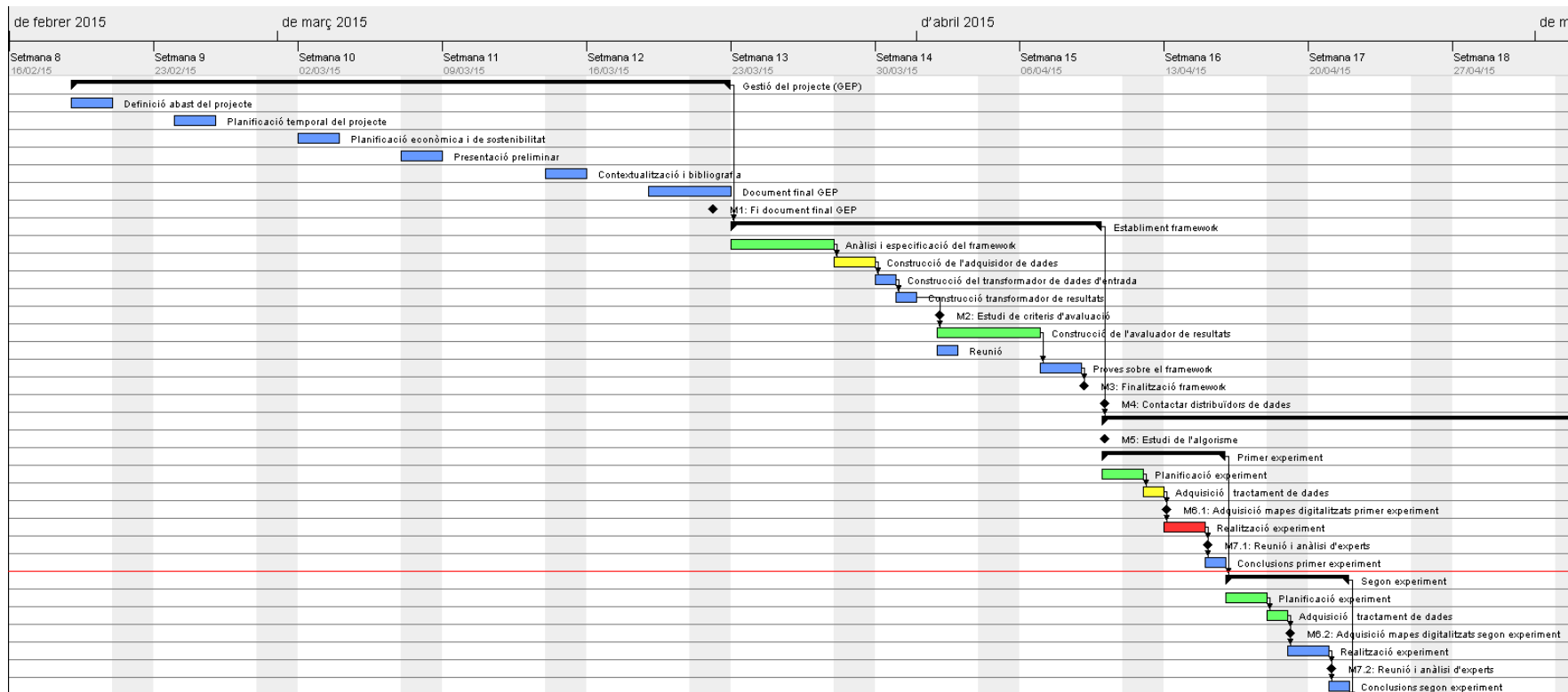
Part textual:

Tasca/<i>Milestone</i>	Inici	Fi	Responsable	Suport
Gestió del projecte (GEP)	19/02/15	22/03/15		
Definició abast del projecte	19/02/15	20/02/15	Cap de projecte	
Planificació temporal del projecte	24/02/15	25/02/15	Cap de projecte	
Planificació econòmica i de sostenibilitat	02/03/15	03/03/15	Cap de projecte	
Presentació preliminar	07/03/15	08/03/15	Cap de projecte	
Contextualització i bibliografia	14/03/15	15/03/15	Cap de projecte	
Document final GEP	19/03/15	22/03/15	Cap de projecte	
M1: Fi document final GEP	22/03/15	22/03/15		
Establiment framework	23/03/15	09/04/15		
Anàlisi i especificació del framework	23/03/15	27/03/15	Dissenyador	
Construcció de l'adquisidor de dades	28/03/15	29/03/15	Programador	Responsable de proves
Construcció del transformador de dades d'entrada	30/03/15	30/03/15	Programador	
Construcció transformador de resultats	31/03/15	31/03/15	Programador	
M2: Estudi de criteris d'avaluació	02/04/15	02/04/15		
Construcció de l'avaluador de resultats	02/04/15	06/04/15	Programador	Responsable de proves
Reunió	02/04/15	02/04/15		
Proves sobre el framework	07/04/15	08/04/15	Responsable de proves	Programador
M3: Finalització framework	09/04/15	09/04/15		
M4: Contactar distribuïdors de dades	10/04/15	10/04/15		
Experiments	10/04/15	13/05/15		
M5: Estudi de l'algorisme	10/04/15	10/04/15		
Primer experiment	10/04/15	15/04/15		
Planificació experiment	10/04/15	11/04/15	Cap de projecte	Dissenyador
Adquisició i tractament de dades	12/04/15	12/04/15	Analista	Programador
M6.1: Adquisició mapes digitalitzats primer experiment	13/04/15	13/04/15		

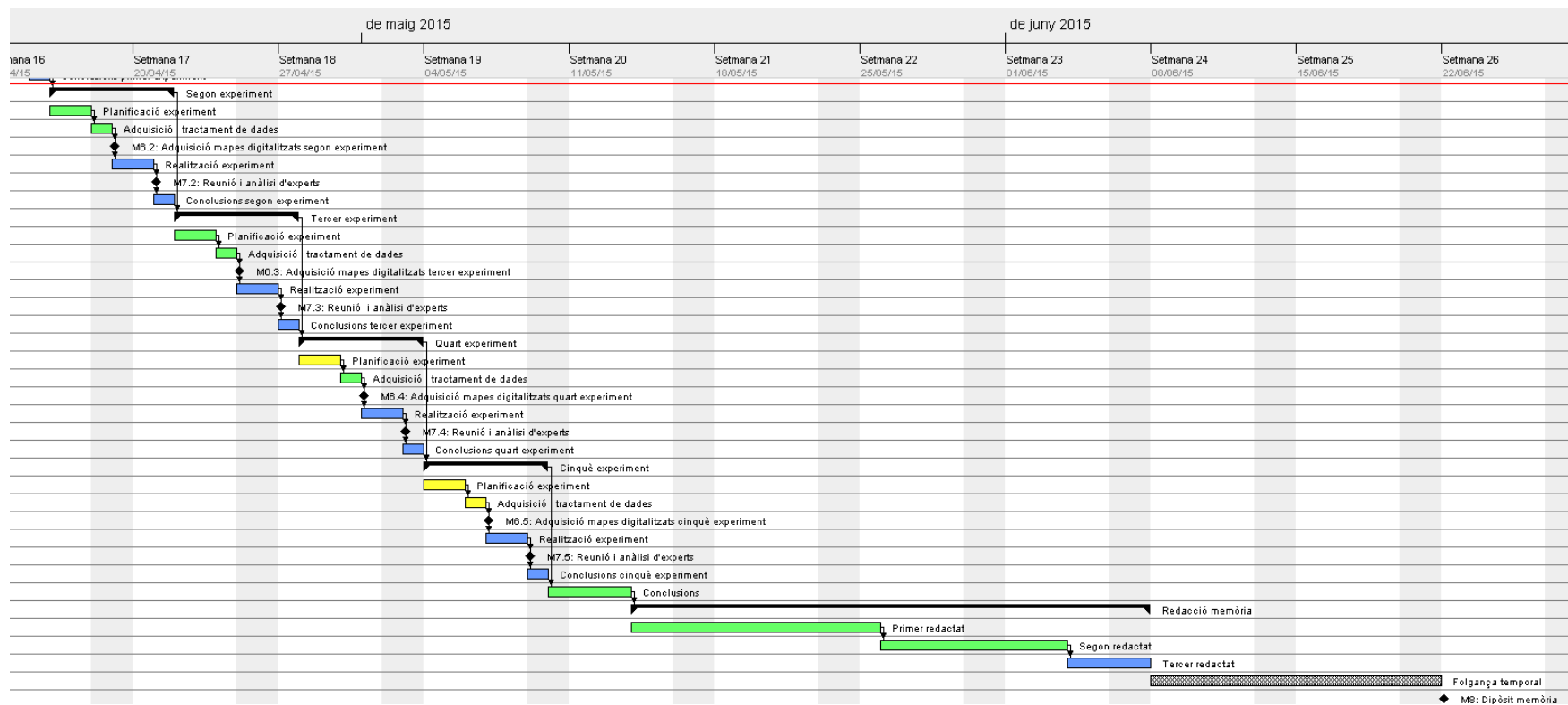
Realització experiment	13/04/15	14/04/15	Analista	
M7.1: Reunió i anàlisi d'experts	15/04/15	15/04/15		
Conclusions primer experiment	15/04/15	15/04/15	Analista	Cap de projecte
Segon experiment	16/04/15	21/04/15		
Planificació experiment	16/04/15	17/04/15	Cap de projecte	Dissenyador
Adquisició i tractament de dades	18/04/15	18/04/15	Analista	Programador
M6.2: Adquisició mapes digitalitzats segon experiment	19/04/15	19/04/15		
Realització experiment	19/04/15	20/04/15	Analista	
M7.2: Reunió i anàlisi d'experts	21/04/15	21/04/15		
Conclusions segon experiment	21/04/15	21/04/15	Analista	Cap de projecte
Tercer experiment	22/04/15	27/04/15		
Planificació experiment	22/04/15	23/04/15	Cap de projecte	Dissenyador
Adquisició i tractament de dades	24/04/15	24/04/15	Analista	Programador
M6.3: Adquisició mapes digitalitzats tercer experiment	25/04/15	25/04/15		
Realització experiment	25/04/15	26/04/15	Analista	
M7.3: Reunió i anàlisi d'experts	27/04/15	27/04/15		
Conclusions tercer experiment	27/04/15	27/04/15	Analista	Cap de projecte
Quart experiment	28/04/15	03/05/15		
Planificació experiment	28/04/15	29/04/15	Cap de projecte	Dissenyador
Adquisició i tractament de dades	30/04/15	30/04/15	Analista	Programador
M6.4: Adquisició mapes digitalitzats quart experiment	01/05/15	01/05/15		
Realització experiment	01/05/15	02/05/15	Analista	
M7.4: Reunió i anàlisi d'experts	03/05/15	03/05/15		
Conclusions quart experiment	03/05/15	03/05/15	Analista	Cap de projecte
Cinquè experiment	04/05/15	09/05/15		
Planificació experiment	04/05/15	05/05/15	Cap de projecte	Dissenyador
Adquisició i tractament de dades	06/05/15	06/05/15	Analista	Programador
M6.5: Adquisició mapes digitalitzats cinquè experiment	07/05/15	07/05/15		
Realització experiment	07/05/15	08/05/15	Analista	

M7.5: Reunió i anàlisi d'experts	09/05/15	09/05/15		
Conclusions cinquè experiment	09/05/15	09/05/15	Analista	Cap de projecte
Conclusions	10/05/15	13/05/15	Cap de projecte	
Redacció memòria	14/05/15	07/06/15		
Primer redactat	14/05/15	25/05/15	Cap de projecte	
Segon redactat	26/05/15	03/06/15	Cap de projecte	
Tercer redactat	04/06/15	07/06/15	Cap de projecte	
Folgança temporal	08/06/15	21/06/15		
M8: Dipòsit memòria	22/06/15	22/06/15		

Part gràfica I:



Part gràfica II:



Milestones o fites

Amb els *milestones* podem tenir objectius clars els quals ens permeten l'assoliment de fites importants i punts de control per valorar el progrés del projecte.

Codi Descripció

- M1* Fi del document final del GEP on s'exposa la organització total del projecte.
- M2* Fi de l'estudi dels criteris d'avaluació per tal de saber quin s'adapta millor a les necessitats del projecte.
- M3* Fi de la creació del *framework* a on es realitzaran els experiments
- M4* Haver contactat amb possibles proveïdors de dades de trajectòries GPS.
- M5* Fi de l'estudi de com funciona l'algorisme. Fita necessària per saber valorar els resultats d'aquest.
- M6* Adquisició de mapes digitalitzats per cada un dels experiments.
- M7* Reunió amb els directors per tal de fer la valoració visual dels resultats.
- M8* Fi del projecte.

Distribució de l'esforç

A partir de les tasques dels projectes podem observar que el percentatge en activitats del projecte és:

Activitat	Contingut	Percentatge
<i>Front-end</i>	Reunions de seguiment	58.14 %
	Gestió del projecte	
	Conclusions experiments	
	Redactat memòria	
	Planificació d'experiments	
Construcció i experiments	Adquisició dades	33.74 %
	Construcció <i>framework</i>	
	Realització experiments	
Proves	Tasques de proves	8.12 %
	Parametrització algorisme	

Recursos materials emprats

Desglossarem a la següent taula els recursos necessaris en les diferents fases del projecte. En aquesta no es mostren els recursos humans, només els de programari i materials. Per veure els recursos humans

Recurs	Tipus	Fases
MATERIAL		
Ordinador	Desenvolupament	totes
Connexió a Internet	Desenvolupament	totes
Materials d'oficina	Desenvolupament	gestió, redacció
PROGRAMARI		
Editor de text L ^A T _E X	Desenvolupament	gestió, redacció
<i>Google Earth</i>	Desenvolupament	experiments
Control de programari Git	Gestió	<i>framework</i> , experiments
GanttProject	Gestió	gestió
Bitbucket	Gestió	<i>framework</i> , experiments
Correu electrònic	Comunicació	totes
Editor de text Vim	Desenvolupament	<i>framework</i> , experiments
Editor d'imatges Inkscape	Desenvolupament	experiments
Python	Desenvolupament	<i>framework</i> , experiments

5.1.1 Valoració d'alternatives i pla d'acció

Per tal d'avaluar el progrés del projecte, es prendran de referència les diferents *milestones* així com la fi de cada una de les fases. A partir d'aquestes referències es valorarà quina desviació temporal existeix entre el temps real i el temps estimat del projecte. A partir de la desviació, es realitzarà el següent pla d'acció

- Si la desviació és lleugera (menor a 7 dies): Hi ha un període final abans de l'entrega de folgança temporal. En cas que la desviació sigui curta es tindrà en compte per analitzar els riscos que aquesta comporta però se seguirà amb la planificació establerta, agafant el temps restant de la folgança.
- Si la desviació és mitjana (entre 7 a 20 dies): S'aplicaran mesures correctives (per exemple, més hores assignades al dia per fer el projecte) i s'ajustarà una nova

planificació reavaluant-ne els riscos i es documentaran els problemes, incidències i imprevistos que hagin sorgit.

- Si la desviació és gran (superior a 20 dies): Es reduirà el nombre d'experiments, escurçant la llargada del projecte i documentant les causes d'aquesta desviació tan important.

Finalment, cal notar que l'endarreriment de les tasques no comporta un augment real en els recursos materials i de programari, ja que aquests estan disponibles en la totalitat del projecte. Tanmateix, sí pot suposar un augment en les hores de treball dels diferents rols, en cas que l'endarreriment vingui ocasionat per subestimar el temps requerit de les tasques. Per aquest motiu, s'ha afegit un pla de contingència en els pressupostos.

5.2 Gestió econòmica (pressupost)

5.2.1 Identificació i estimació dels costos

El recompte d'hores s'ha fet per jornades diàries de 3 h i mitja, en el període de 133 dies de duració del projecte. El projecte es durà a terme, per tant, en aproximadament 465.5 hores de treball.

Recursos humans Malgrat el projecte ha sigut realitzat per una sola persona, en el següent pressupost es mostra una aproximació del preu dels recursos humans mitjançant rols i el sou d'aquests, tarifant per hores les feines de cada un dels rols.

Rol	Hores estimades	Remuneració ¹	Cost estimat
Cap de projecte	190	18.83 €/h	3577.70 €
Analista	153	16.56 €/h	2533.68 €
Dissenyador	33	15.28 €/h	504.24 €
Programador	62	9.04 €/h	560.48 €
Responsable de proves	28	9.42 €/h	263.76 €
Total	466		7439.86 €

Recursos materials Els recursos *software* del projecte seran de codi lliure o gratuïts (*freeware*), per tant no es comptabilitzen en el pressupost. Pel càlcul d'amortització s'ha considerat la vida útil del producte respecte el període del projecte de 133 dies.

Producte	Temps de vida útil	Cost	Cost amortitzat
Ordinador personal	5 anys	1000.00 €	72.88 €
Total			72.88 €

Altres costos

Concepte	Cost estimat
Electricitat (0.26 kW en 466 h a 0.125 €/kWh) ²	15.15 €
Internet (5 mesos)	150.00 €
Material d'oficina (inclou impressió memòria)	120.00 €
Total	285.15 €

Desglossament temporal dels pressupostos Els recursos invertits en cada fase del diagrama de Gantt es poden desglossar de la següent manera:

Fase	Hores	Cost
GESTIÓ DEL PROJECTE		
<i>Recursos humans</i>		
Cap de projecte	62 h	1167.46 €
Subtotal	62 h	1167.46 €

¹Segons dades de *Page personnel*, disponibles a http://www.pagepersonnel.es/productsApp_pp_es/Estudios%20Remuneracion/er_tecnologia.pdf.

²Segons dades de <http://tarifasgasluz.com/faq/precio-kwh/espana>.

ESTABLIMENT FRAMEWORK		
<i>Recursos humans</i>		
Dissenyador	4 h	61.12 €
Programador	60 h	542.40 €
Responsable de proves	28 h	263.76 €
Subtotal	92 h	867.28 €
EXPERIMENTS		
<i>Recursos humans</i>		
Cap de projecte	5 h	94.15 €
Analista	153 h	2533.68 €
Dissenyador	29 h	443.12 €
Programador	2 h	18.08 €
Subtotal	189 h	3089.03 €
REDACCIÓ MEMÒRIA		
<i>Recursos humans</i>		
Cap de projecte	123 h	2316.03 €
Subtotal	123 h	2316.03 €
RECURSOS MATERIALS I ALTRES		
<i>Amortitzacions</i>		72.88 €
<i>Altres costos</i>		285.15 €
Subtotal		358.03 €
Total		7797.89 €

Pressupost total Com que hi poden haver imprevistos per diferents riscos com l'augment de duració de les tasques, la necessitat d'adquisició de dades de trajectòries GPS privades, l'adquisició de mapes digitalitzats privats, . . . , s'ha inclòs un marge del 10% sobre el total del projecte.

Tipus cost	Cost estimat
Recursos humans	7439.86 €
Recursos materials (amort.)	72.88 €
Altres costos	285.15 €
<i>Subotal</i>	<i>7797.89 €</i>
Contingència	10 %
Total	8577.68 €

Finalment, cal dir que al ser un projecte acadèmic, aquest pressupost és fictici.

5.2.2 Control de gestió

Es mantindrà un registre de les hores de feina realitzades per tal de veure si es sobrepassa el número d'hores planificat. Aquestes dades es revisaran juntament amb la planificació temporal, especialment al final de cada experiment i fase.

Si el cost real fos major al pressupostat, s'assignarà la partida de contingència a efecte de cobrir la diferència. En qualsevol cas, es recolliran en una taula els costos estimats, els costos reals i la desviació en percentatge a efectes de documentar i analitzar d'on han sortit les desviacions.

Les desviacions a considerar per tal de saber perquè i de quin import de desviació parlem en cada cas són:

En preu

- Desviament realització d'una tasca = (cost estimat - cost real) * consum hores real
- Desviament d'un recurs = (cost estimat - cost real) * consum real

En consum

- Desviament realització d'una tasca = (consum hores estimat - consum hores real) * cost estimat
- Desviament d'un recurs = (consum hores estimat - consum hores real) * cost estimat

Totals

- Desviament per volum en vendes = (vendes reals - vendes pressupostades) * marge estimat
- Desviament per marge en vendes = (marge real - marge estimat) * vendes reals
- Desviament total en mà d'obra = cost estimat - cost total
- Desviament total en recursos = cost estimat - cost total
- Desviament total en matèria primera = cost estimat - cost total
- Desviament total en vendes = total vendes - total vendes reals
- Desviament total en costos fixos = cost estimat - cost total

5.3 Sostenibilitat i compromís social**5.3.1 Matriu de sostenibilitat**

	<i>Econòmica</i>	<i>Social</i>	<i>Mediambiental</i>	<i>Total</i>
<i>Planificació</i>	4	7	9	21

5.3.2 Sostenibilitat econòmica

S'ha realitzat una avaluació de costos completa, així com s'ha inclòs un marge de contingència. Donat que el projecte està orientat a la recerca, no representa un producte que pugui ser venut directament, per tant, el sector de mercat és molt reduït i enfocat a departaments d'I+D, dificultant la viabilitat econòmica del projecte. Donat que la gran part del preu ve de recursos humans (95.37%), el preu final es podria veure reduït

notablement, per exemple, agrupant rols en una sola persona amb un sou inferior al comptabilitzat.

Finalment, a causa del mercat reduït que té, es considera un 4 en sostenibilitat econòmica.

5.3.3 Sostenibilitat social

Aquest projecte afavorirà la investigació en la generació automàtica de mapes en llocs de senderisme. El benefici que s'obtindrà a partir d'aquí és força variat. Per una banda es podran millorar els serveis de rescat de persones en aquestes zones al disposar de mapes sempre actualitzats. Per un altre banda, les persones també podran beneficiar-se a l'hora de planificar les seves activitats d'oci a l'aire lliure com serien el senderisme, el BTT, esport en muntanya, . . . , millorant així la qualitat de vida de les persones.

Tanmateix, els algorismes requereixen de trajectòries GPS per funcionar (i com més actuals millor). Aquest requeriment planteja el problema que un col·lectiu de gent hauria d'estar disposada d'enregistrar les seves posicions quan fan activitats en rutes de senderisme, posant sobre la taula el problema de la privacitat de les dades.

Considerant els avantatges i els inconvenients socials, finalment s'ha considerat un 7 en la sostenibilitat social d'aquest projecte.

5.3.4 Sostenibilitat mediambiental

L'elaboració del projecte requerirà d'electricitat així com d'un ordinador i material d'oficina. Tenint en compte que la major part del material d'oficina pot ser reciclable aquest no esdevé un problema. El total de CO₂ generat per kWh s'estima en 0.30 kg/kWh segons dades oficials³ del govern espanyol. Per tant, el projecte generaria una quantitat de 36.35 kg de CO₂. Finalment, cal tenir en consideració que l'ordinador requerit acabarà sent deixalla electrònica difícilment reciclable.

Considerant els factors anteriors, entenent que la quantitat de CO₂ emès és moderat, i que el temps en què s'utilitza l'ordinador és tan sols una part de la seva vida útil, s'ha puntuat la sostenibilitat mediambiental amb un 9.

³http://www.minetur.gob.es/energia/desarrollo/EficienciaEnergetica/RITE/propuestas/Documents/2014_03_03_Factores_de_emision_CO2_y_Factores_de_paso_Efinal_Eprimaria_V.pdf

5.4 Relació amb l'especialitat de Computació

5.4.1 Assignatures relacionades

- A: Per tal d'entendre l'algorisme seran importants les bases d'algorísmia oferta per aquesta assignatura. Per exemple, per entendre costos asimptòtics, el paradigma incremental, algorismes sobre grafs i grafs geomètrics,...
- LP i CL: es realitzaran diversos *parsers* per tal de transformar entre els diferents tipus de dades. A més, es faran un conjunt de *scripts*.
- CN: Servirà per entendre possibles errors d'aproximació en el càlcul numèric amb l'aritmètica de coma flotant.

5.4.2 Adequació amb l'especialitat

- Aquest projecte té una forta component d'investigació.
- En aquest projecte hi ha l'avaluació i comprensió d'algorismes, molt comú en l'especialitat.
- Els algorismes tenen una forta component matemàtica i geomètrica.
- L'anàlisi de la tipologia de dades així com l'establiment de criteris per l'avaluació de models matemàtics són pròpies de l'especialitat.
- Té una forta component de teoria de grafs, de fet, el resultat és un graf geomètric representant la xarxa de camins.
- Es necessiten reconeixadors (*parsers*) de llenguatges de marques (especificacions de xml) pel pre-tractament i post-tractament de les dades.

5.4.3 Competències relacionades

CCO1.1	Avaluar la complexitat computacional d'un problema, conèixer estratègies algorísmiques que puguin dur a la seva resolució, i recomanar, desenvolupar i implementar la que garanteixi el millor rendiment d'acord amb els requisits establerts.	En profunditat
---------------	--	----------------

CCO2.1	Demostrar coneixement dels fonaments, dels paradigmes i de les tècniques pròpies dels sistemes intel·ligents, i analitzar, dissenyar i construir sistemes, serveis i aplicacions informàtiques que utilitzin aquestes tècniques en qualsevol àmbit d'aplicació.	Bastant
CCO2.2	Capacitat per a adquirir, obtenir, formalitzar i representar el coneixement humà d'una forma computable per a la resolució de problemes mitjançant un sistema informàtic en qualsevol àmbit d'aplicació, particularment en els que estan relacionats amb aspectes de computació, percepció i actuació en ambients o entorns intel·ligents.	Una mica

5.5 Llicències i normatives

- Els algorismes analitzats tenen llicències de codi obert i són accessibles, per tant, es poden utilitzar pel propòsit d'aquest treball.
- El programa Google Earth en permet un ús personal lliure de càrrecs i, sota la legislació americana (que és sota la qual es regeix els termes i condicions d'ús d'aquest programari), l'ús que se n'ha fet entra dins del "fair use" ja que el treball entra dins de docència i investigació.
- Els mapes de OpenStreetMap usats per a la generació dels *ground truth* així com en la visualització són de lliure ús i distribució.
- La pàgina web Wikilocs permet la descàrrega de rutes GPS per a ús personal i, sota la legislació espanyola, es contempla l'ús que se'n fa de les trajectòries com a excepció i límit del copyright, ja que es fan servir per docència i investigació.

6 Conclusions

Primer s'expliquen les conclusions dels resultats (Secció 6.1) i, a continuació, quina feina es pot realitzar en un futur (Secció 6.2). Finalment, a títol personal, faig una valoració transversal d'aquest treball (Secció 6.3).

6.1 Conclusions dels resultats

En aquest treball s'ha fet una anàlisi innovadora de dos dels algorismes més recents en la generació de mapes de carrers i carreteres de ciutat al ser aplicats al senderisme. A més a més, també s'han identificat per primera vegada les principals diferències entre trajectòries GPS provinents de vehicles per ciutat, de senderistes per terreny pla i de senderistes per muntanyes. Finalment, s'ha aplicat i justificat una metodologia per a la correcció de *ground-truth*. No obstant, les correccions han resultat ser insuficient per aconseguir l'objectiu inicial de poder comparar les mesures de qualitat obtingudes en els mapes generats de ciutat i els de senderisme, a causa de la poca significació d'aquests darrers per la manca d'un bon *ground-truth*.

Ahmed i Wenk [4] és un algorisme senzill però incapaç de treballar sota condicions de soroll i, a més, no utilitza les marques de temps de les trajectòries, fet que l'incapacita de treballar en zones de geometria complexa on no n'hi ha prou amb la «proximitat». En zones pròpies del senderisme de muntanya conviuen *outliers*, soroll i camins pròxims uns als altres. La falta de mètodes capaços de distingir aquestes tres situacions juntament amb la manca de tècniques que considerin les mitjanes entre les trajectòries, fan que aquest algorisme sigui incapaç d'abordar satisfactòriament el problema de la generació de mapes en senderisme de muntanya. Malgrat tot, la seva senzillesa i comportament estable el fan candidat a ser la base d'algorismes més complexos capaços de cobrir les seves mancances.

En l'extrem oposat, Karagiorgou i Pfoser [31] és un algorisme complex que intenta extreure informacions no trivials de les trajectòries, com són les interseccions del mapa.

No obstant, aquesta complexitat és la causa de la seva rigidesa i incapacitat d'adaptació a altres tipus de trajectòries. Concretament, a fi de tractar millor les trajectòries de ciutat, aquest algorisme integra coneixements d'aquest domini específic, com serien la reducció de velocitat quan es produeix un gir i el model de 8 tipus de gir. Tanmateix, aquests coneixements no són aplicables al senderisme i, per tant, la seva integració és perjudicial per la qualitat dels resultats. A més, aquesta complexitat significa un cost computacional més elevat.

Amb relació a les zones de senderisme, a partir dels resultats hem pogut constatar que els mapes generats en senderisme de muntanya són de força menor qualitat que els generats en senderisme de terreny pla. En canvi, aquesta diferència no és tan important quan comparem els mapes generats en senderisme de terreny pla amb els generats en una ciutat. Aquest fet indica que la «bondat» d'un conjunt de dades per als algorismes que analitzem depèn més de factors com el soroll de les trajectòries, la rectitud dels camins i l'obertura dels angles en les interseccions (diferències principals entre senderisme de muntanya i de terreny pla), que no pas aspectes com la velocitat, el soroll angular degut a la baixa velocitat o el tipus de moviment d'una persona envers un vehicle (diferències principals entre ciutat i senderisme de terreny pla).

Més enllà dels algorismes i les trajectòries, treballar en zones de senderisme limita la cobertura i qualitat dels *ground-truth* disponibles. Malgrat ser aquesta una de les motivacions per la recerca de generació de mapes en senderisme, a la vegada en dificulta l'anàlisi ja que tres de les quatre mesures de qualitat proposades a Ahmed et al. [5] es disparden quan el mapa generat conté camins que el *ground-truth* no cobreix. Com a conseqüència, sense un *ground-truth* amb cobertura completa del terreny no s'ha pogut obtenir una mesura fiable de la qualitat dels resultats i, per tant, no s'han pogut comparar amb les mesures obtingudes en la ciutat.

A tall de conclusió, aquest treball consisteix en una anàlisi de les diferències entre la generació de mapes de ciutat i de senderisme. A causa de la maduresa en la investigació de la generació de mapes de ciutat, creiem que aquesta anàlisi obre les portes a un nombre elevat de succedànies investigacions, tal i com plantegem a la següent secció de feina a realitzar en un futur (Secció 6.2).

6.2 Feina a realitzar en un futur

A partir dels resultats experimentals obtinguts en aquest treball, hem identificat quatre direccions per on seguir una futura recerca dins del context de la generació de mapes de senderisme. Aquestes quatre direccions consisteixen en *analitzar* més algorismes de generació de mapes de ciutat, *adaptar* algorismes perquè treballin millor amb dades de senderisme, *crear* nous algorismes pensats en la generació de dades de senderisme i *descobrir* mesures de qualitat més adaptades als *ground-truth* propis del senderisme.

La direcció d'*analitzar* més algorismes consisteix a ampliar els esforços d'aquest treball. Concretament, Ahmed i Wenk [4] i Karagiorgou i Pfoser [31] són algorismes d'inserció incremental de trajectòries i d'unió d'interseccions, respectivament. En l'àmbit d'aquest treball no s'ha analitzat cap algorisme de creació de clústers de punts com Cao i Krumm [14] i Edelkamp i Schrödl [19], els quals aprofundirien els resultats analitzant l'adequació de l'enfocament dels clústers de punts en la generació de mapes de senderisme.

Un pas més enllà a simplement analitzar els algorismes és *adaptar*-los al senderisme. Algunes d'aquestes adaptacions ja han sigut mencionades en aquest treball com és el cas de treballar en tres dimensions o permetre una parametrització variable segons la zona del mapa.

Una altra direcció d'investigació interessant és la *creació* d'un algorisme pensat per generar mapes de senderisme. Aquesta creació, però, no partiria de zero ja que en aquest treball s'ha realitzat un estudi de les característiques de les trajectòries així com l'adequació al senderisme d'algunes de les idees de Ahmed i Wenk [4] i Karagiorgou i Pfoser [31].

Finalment, en aquest treball s'ha mostrat empíricament com les mesures de qualitat actuals no toleren que el *ground-truth* cobreixi només un subconjunt dels camins del mapa generat. Donat que en el senderisme és difícil trobar bons *ground-truth* que cobreixin tot el terreny, per tal de poder fer una comparació quantitativa de la qualitat s'han de trobar noves mesures de qualitat.

6.3 Conclusions acadèmiques i personals

Aquest treball final de grau m'ha permès treballar amb un tema de recerca actual. Com a tal, he pogut viure de primera mà la satisfacció de realitzar un treball únic, on els resultats obtinguts acabaran formant part del coneixement científic de la nostra societat. A més, he tingut la sort de poder experimentar com és treballar en un equip d'investigació format per jo, els directors i un estudiant de la Facultat de Matemàtiques i Estadística cursant un treball final de grau paral·lel al meu [32]. Formar part d'aquest equip interdisciplinari m'ha permès estar en un ambient enriquidor on es presentaven punts de vista diferents que han donat peu a discussions fructíferes. Per tant, només puc reafirmar que la meva valoració global del projecte ha sigut completament satisfactòria.

Referències

- [1] Mridul Aanjaneya, Frederic Chazal, Daniel Chen, Marc Glisse, Leonidas J. Guibas i Dmitriy Morozov. «Metric Graph Reconstruction from Noisy Data». A: *Proceedings of the Twenty-seventh Annual Symposium on Computational Geometry*. SoCG '11. Paris, France: ACM, 2011, pàg. 37 - 46. ISBN: 978-1-4503-0682-9. DOI: [10.1145/1998196.1998203](https://doi.org/10.1145/1998196.1998203). URL: <http://doi.acm.org/10.1145/1998196.1998203>.
- [2] Gabriel Agamennoni, Juan I. Nieto i Eduardo M. Nebot. «Robust Inference of Principal Road Paths for Intelligent Transportation Systems». A: *Intelligent Transportation Systems, IEEE Transactions on* 12.1 (2011), pàg. 298 - 308. ISSN: 1524-9050. DOI: [10.1109/TITS.2010.2069097](https://doi.org/10.1109/TITS.2010.2069097).
- [3] Mahmuda Ahmed, Kyle S. Hickmann i Carola Wenk. «Path-Based Distance for Street Map Comparison». A: *CoRR* abs/1309.6131 (2013). URL: <http://arxiv.org/abs/1309.6131>.
- [4] Mahmuda Ahmed i Carola Wenk. «Constructing Street Networks from GPS Trajectories». A: *Proceedings of the 20th Annual European Conference on Algorithms*. ESA'12. Ljubljana, Slovenia: Springer-Verlag, 2012, pàg. 60 - 71. ISBN: 978-3-642-33089-6. DOI: [10.1007/978-3-642-33090-2_7](https://doi.org/10.1007/978-3-642-33090-2_7). URL: http://dx.doi.org/10.1007/978-3-642-33090-2_7.
- [5] Mahmuda Ahmed, Sophia Karagiorgou, Dieter Pfoser i Carola Wenk. «A comparison and evaluation of map construction algorithms using vehicle tracking data». English. A: *GeoInformatica* 19.3 (2015), pàg. 601 - 632. ISSN: 1384-6175. DOI: [10.1007/s10707-014-0222-6](https://doi.org/10.1007/s10707-014-0222-6). URL: <http://dx.doi.org/10.1007/s10707-014-0222-6>.
- [6] Helmut Alt i Michael Godau. «Computing the Fréchet distance between two polygonal curves». A: *International Journal of Computational Geometry & Applications* 05.01n02 (1995), pàg. 75 - 91. DOI: [10.1142/S0218195995000064](https://doi.org/10.1142/S0218195995000064).
- [7] Helmut Alt i Leonidas Guibas. *Discrete Geometric Shapes: Matching, Interpolation, and Approximation: A Survey*. Handbook of Computational Geometry, 1996, pàg. 121 - 154.

-
- [8] Helmut Alt, Alon Efrat, Günter Rote i Carola Wenk. «Matching Planar Maps». A: *J. Algorithms* 49.2 (nov. de 2003), pàg. 262-283. ISSN: 0196-6774. DOI: [10.1016/S0196-6774\(03\)00085-3](https://doi.org/10.1016/S0196-6774(03)00085-3). URL: [http://dx.doi.org/10.1016/S0196-6774\(03\)00085-3](http://dx.doi.org/10.1016/S0196-6774(03)00085-3).
- [9] Diputació de Barcelona. *Carotgrafia Parc Natural del Montseny*. URL: <http://parcs.diba.cat/web/montseny/mapes> (cons. 1-6-2015).
- [10] Diputació de Barcelona. *Cartografia Parc Natural del Garraf*. URL: <http://parcs.diba.cat/web/garraf/cartografia> (cons. 1-6-2015).
- [11] James Biagioni i Jakob Eriksson. «Inferring road maps from GPS traces: Survey and comparative evaluation». A: *In Transportation Research Board, 91st Annual*. 2012, pàg. 61-71.
- [12] James Biagioni i Jakob Eriksson. «Map Inference in the Face of Noise and Disparity». A: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. SIGSPATIAL '12. Redondo Beach, California: ACM, 2012, pàg. 79-88. ISBN: 978-1-4503-1691-0. DOI: [10.1145/2424321.2424333](https://doi.org/10.1145/2424321.2424333). URL: <http://doi.acm.org/10.1145/2424321.2424333>.
- [13] R. Bruntrup, S. Edelkamp, S. Jabbar i B. Scholz. «Incremental map generation with GPS traces». A: *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*. 2005, pàg. 574-579. DOI: [10.1109/ITSC.2005.1520084](https://doi.org/10.1109/ITSC.2005.1520084).
- [14] Lili Cao i John Krumm. «From GPS Traces to a Routable Road Map». A: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. GIS '09. Seattle, Washington: ACM, 2009, pàg. 3-12. ISBN: 978-1-60558-649-6. DOI: [10.1145/1653771.1653776](https://doi.org/10.1145/1653771.1653776). URL: <http://doi.acm.org/10.1145/1653771.1653776>.
- [15] Chen Chen i Yinhang Cheng. «Roads Digital Map Generation with Multi-track GPS Data». A: *Proceedings of the 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing - Volume 01*. ETTANDGRS '08. Washington, DC, USA: IEEE Computer Society, 2008, pàg. 508-511. ISBN: 978-0-7695-3563-0. DOI: [10.1109/ETTandGRS.2008.70](https://doi.org/10.1109/ETTandGRS.2008.70). URL: <http://dx.doi.org/10.1109/ETTandGRS.2008.70>.
- [16] Daniel Chen, Leonidas J. Guibas, John Hershberger i Jian Sun. «Road Network Reconstruction for Organizing Paths». A: *Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '10. Austin, Texas: Society

- for Industrial i Applied Mathematics, 2010, pàg. 1309 - 1320. ISBN: 978-0-898716-98-6. URL: <http://dl.acm.org/citation.cfm?id=1873601.1873706>.
- [17] Microsoft Corporation. *Bing Maps*. 2010. URL: <https://www.bing.com/maps/> (cons. 1-6-2015).
- [18] Jonathan J. Davies, Alastair R. Beresford i Andy Hopper. «Scalable, Distributed, Real-Time Map Generation». A: *IEEE Pervasive Computing* 5.4 (oct. de 2006), pàg. 47 - 54. ISSN: 1536-1268. DOI: [10.1109/MPRV.2006.83](https://doi.org/10.1109/MPRV.2006.83). URL: <http://dx.doi.org/10.1109/MPRV.2006.83>.
- [19] Stefan Edelkamp i Stefan Schrödl. «Computer Science in Perspective». A: ed. de Rolf Klein, Hans-Werner Six i Lutz Wegner. New York, NY, USA: Springer-Verlag New York, Inc., 2003. Cap. Route Planning and Map Inference with Global Positioning Traces, pàg. 128 - 151. ISBN: 3-540-00579-X. URL: <http://dl.acm.org/citation.cfm?id=865449.865460>.
- [20] James J. Spilker Jr.; Penina Axelrad; Bradford W. Parkinson; Per Enge. *Global Positioning System: Theory and Applications, Volume I*. 1996. DOI: [10.2514/4.866388](https://doi.org/10.2514/4.866388).
- [21] Alireza Fathi i John Krumm. «Detecting Road Intersections from GPS Traces». A: *Proceedings of the 6th International Conference on Geographic Information Science*. GIScience'10. Zurich, Switzerland: Springer-Verlag, 2010, pàg. 56 - 69. ISBN: 3-642-15299-6, 978-3-642-15299-3. URL: <http://dl.acm.org/citation.cfm?id=1887961.1887966>.
- [22] OpenStreetMap Foundation. *OpenStreetMap*. 2006. URL: <https://www.openstreetmap.org> (cons. 1-6-2015).
- [23] OpenStreetMap Foundation. *OSM XML*. 2006. URL: http://wiki.openstreetmap.org/wiki/OSM_XML (cons. 1-6-2015).
- [24] Xiaoyin Ge, Issam I. Safa, Mikhail Belkin i Yusu Wang. «Data Skeletonization via Reeb Graphs». A: *Advances in Neural Information Processing Systems 24*. Ed. de J. Shawe-taylor, R.s. Zemel, P. Bartlett, F.c.n. Pereira i K.q. Weinberger. 2011, pàg. 837 - 845. URL: http://books.nips.cc/papers/files/nips24/NIPS2011_0559.pdf.
- [25] Google. *Keyhole Markup Language*. 2008. URL: <https://developers.google.com/kml> (cons. 1-6-2015).
- [26] *GPX The GPS Exchange Format*. 2002. URL: <http://www.topografix.com/gpx.asp> (cons. 1-6-2015).

- [27] Leonidas J. Guibas, John E. Hersherberger, Joseph S.B. Mitchell i Jack Scott Snoeyink. «Approximating polygons and subdivisions with minimum link paths». English. A: *ISA '91 Algorithms*. Ed. de Wen-Lian Hsu i R.C.T. Lee. Vol. 557. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1991, pàg. 151-162. ISBN: 978-3-540-54945-1. DOI: [10.1007/3-540-54945-5_59](https://doi.org/10.1007/3-540-54945-5_59). URL: http://dx.doi.org/10.1007/3-540-54945-5_59.
- [28] Tao Guo, K. Iwamura i M. Koga. «Towards high accuracy road maps generation from massive GPS Traces data». A: *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*. 2007, pàg. 667-670. DOI: [10.1109/IGARSS.2007.4422884](https://doi.org/10.1109/IGARSS.2007.4422884).
- [29] Dirk Stöcker Immanuel Scholz. *JOSM*. 2015. URL: <https://josm.openstreetmap.de/> (cons. 1-6-2015).
- [30] Sera Jang, Taehwan Kim i Seunghwa Lee. «Map generation system with lightweight GPS trace data». A: *Advanced Communication Technology (ICACT), 2010 The 12th International Conference on*. Vol. 2. 2010, pàg. 1489-1493.
- [31] Sophia Karagiorgou i Dieter Pfoser. «On Vehicle Tracking Data-based Road Network Generation». A: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. SIGSPATIAL '12. Redondo Beach, California: ACM, 2012, pàg. 89-98. ISBN: 978-1-4503-1691-0. DOI: [10.1145/2424321.2424334](https://doi.org/10.1145/2424321.2424334). URL: <http://doi.acm.org/10.1145/2424321.2424334>.
- [32] Enle Lin. «Generació de mapes de senderisme a partir de trajectòries GPS». Tesis grau. Universitat Politècnica de Catalunya, 2015. Disp. en breu.
- [33] Xuemei Liu, James Biagioni, Jakob Eriksson, Yin Wang, George Forman i Yanmin Zhu. «Mining Large-scale, Sparse GPS Traces for Map Inference: Comparison of Approaches». A: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '12. Beijing, China: ACM, 2012, pàg. 669-677. ISBN: 978-1-4503-1462-6. DOI: [10.1145/2339530.2339637](https://doi.org/10.1145/2339530.2339637). URL: <http://doi.acm.org/10.1145/2339530.2339637>.
- [34] Juliane Mondzsch i Monika Sester. «Quality Analysis of OpenStreetMap Data Based on Application Needs». A: *Cartographica* 46.2 (2011), pàg. 115-125. DOI: [10.3138/cart0.46.2.115](https://doi.org/10.3138/cart0.46.2.115). eprint: <http://dx.doi.org/10.3138/cart0.46.2.115>. URL: <http://dx.doi.org/10.3138/cart0.46.2.115>.

- [35] Brian Niehofer, Ralf Burda, Christian Wietfeld, Franziskus Bauer i Oliver Lueert. «GPS Community Map Generation for Enhanced Routing Methods Based on Trace-Collection by Mobile Phones». A: *Proceedings of the 2009 First International Conference on Advances in Satellite and Space Communications*. SPACOMM '09. Washington, DC, USA: IEEE Computer Society, 2009, pàg. 156-161. ISBN: 978-0-7695-3694-1. DOI: [10.1109/.30](https://doi.org/10.1109/.30). URL: <http://dx.doi.org/10.1109/.30>.
- [36] International Association of Oil & Gas Producers. «Geodetic awareness guidance note». A: *Surveying & Positioning Guidance*. 2008. URL: <http://www.iogp.org/pubs/373-01.pdf>.
- [37] Wikiloc Outdoor. *Wikiloc*. 2006. URL: <http://www.wikiloc.com/wikiloc/home.do> (cons. 1-6-2015).
- [38] Stefan Schroedl, Kiri Wagstaff, Seth Rogers, Pat Langley i Christopher Wilson. «Mining GPS Traces for Map Refinement». A: *Data Min. Knowl. Discov.* 9.1 (jul. de 2004), pàg. 59-87. ISSN: 1384-5810. DOI: [10.1023/B:DAMI.0000026904.74892.89](https://doi.org/10.1023/B:DAMI.0000026904.74892.89). URL: <http://dx.doi.org/10.1023/B:DAMI.0000026904.74892.89>.
- [39] Wenhuan Shi, Shuhan Shen i Yuncai Liu. «Automatic generation of road network map from massive GPS, vehicle trajectories». A: *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*. 2009, pàg. 1-6. DOI: [10.1109/ITSC.2009.5309871](https://doi.org/10.1109/ITSC.2009.5309871).
- [40] Stewart Worrall i Eduardo Nebot. «Automated Process for Generating Digitised Maps through». A: *GPS Data Compression, in Australasian Conference on Robotics and Automation*. 2007.
- [41] Jianyang Zheng, Yinhai Wang i Nancy L. Nihan. «Quantitative evaluation of GPS performance under forest canopies». A: *Networking, Sensing and Control, 2005. Proceedings. 2005 IEEE*. Mar. de 2005, pàg. 777-782. DOI: [10.1109/ICNSC.2005.1461289](https://doi.org/10.1109/ICNSC.2005.1461289).