

Indice

Introduzione	3
1. Il rilevamento della posizione e la mappa digitale	12
1.1 La tecnologia GPS.....	13
1.2 Il segmento spaziale	16
1.3 Gli errori di misura del GPS.....	20
1.4 Misure di distanza	23
1.5 La mappa digitale	25
1.6 Tecniche di proiezione	26
1.7 La proiezione di Gauss-Boaga	30
1.8 Errori nella creazione della mappa digitale.....	33
2. Gli algoritmi di <i>map matching</i>	36
2.1 L'identificazione del percorso.....	36
2.2 Tipologie dell'approccio geometrico	41
2.3 L'approccio topologico	46
2.4 Le procedure avanzate per il <i>map matching</i>	48
2.5 La valutazione dei risultati	54
3. L'implementazione di alcune tipologie di <i>map matching</i>	57
3.1 Un esempio di algoritmo topologico.....	58
3.1.1 La ricerca della similarità e dell'orientamento	59
3.1.2 Una procedura in presenza di incroci	64
3.1.3 La stima delle coordinate.....	69
3.1.4 La ricerca del percorso più breve.....	71
3.2 L'approccio probabilistico	77
3.2.1 Un esempio di algoritmo probabilistico.....	79

3.2.2	La regione dell' errore	84
3.3	Applicazioni del modello <i>fuzzy</i> al <i>map matching</i>	88
3.3.1	Un modello fuzzy per dati affetti da "effetto canyon"	93
3.3.2	Una variante al modello fuzzy logic.....	95
4.	Una proposta operativa: la penalizzazione e il <i>windowing</i>	97
4.1	Il costo dello scostamento.....	97
4.2	Il <i>windowing</i>	102
4.3	La definizione delle penalizzazioni	105
4.4	I diversi passi della procedura	107
4.5	Un esempio con dati sperimentali.....	109
4.6	Considerazioni conclusive e possibili sviluppi di ricerca	115
	Conclusioni.....	118
	Appendice A.....	121
	Appendice B	126
	Appendice C	129
	Appendice D.....	138
	Appendice E	142
	Bibliografia.....	151

Introduzione

Nell'ambito delle metodologie per la pianificazione dei sistemi di trasporto, i modelli di viaggio basati sull'analisi dei comportamenti degli individui sono ancora attualmente quelli che risultano i più adatti a rappresentare il fenomeno della mobilità ed i più efficaci per la previsione della domanda e la valutazione degli interventi. In particolare, l'analisi del meccanismo, ovvero degli attributi ed in generale degli aspetti che influenzano e governano il comportamento degli individui nella scelta del percorso (processo comportamentale di scelta spaziale), è cruciale per la quantificazione dei flussi di traffico sull'offerta di trasporto disponibile (archi stradali) e di conseguenza per il dimensionamento delle reti di trasporto e per incentivare politiche di trasporto alternative.

La scelta del percorso è caratterizzata da un processo decisionale complesso e dinamico che per essere analizzato richiede la conoscenza di informazioni relative alle caratteristiche dello spostamento (motivo dello spostamento, ora dello spostamento, origine e destinazione, modo utilizzato, distanza e costi, *etc.*), l'identificazione delle alternative disponibili in relazione all'origine e alla destinazione dello spostamento (un elemento di complessità nella scelta del percorso risiede proprio nella natura non discreta delle alternative di percorso disponibili) e informazioni relative alla propensione di scelta dell'individuo (conoscenza spaziale della rete, esperienza, abitudini e preferenze). L'acquisizione di informazioni accurate dal punto di vista spaziale e temporale, inoltre, è fondamentale non solo per l'analisi approfondita del meccanismo di scelta,

ma ancora di più per la costruzione di un modello di scelta del percorso affidabile in fase di predizione.

L'introduzione di nuove tecnologie di informazione e comunicazione che ha caratterizzato gli ultimi 10 anni, ha fornito un contributo eccezionale al miglioramento della qualità dei dati di attività e viaggio, e in particolare dei percorsi seguiti. Se infatti, tradizionalmente, l'acquisizione di dati spaziali di viaggio accurati e' stata sempre estremamente difficile, in quanto basata prevalentemente sulla capacità del rispondente di ricordare i percorsi seguiti, i tragitti anche brevi etc., l'introduzione del Global Positioning System (GPS) nel campo della ricerca sui trasporti, ha consentito l'ottenimento di dati di localizzazione precisi, il tracciamento anche degli spostamenti di breve distanza, e la rilevazione precisa del percorso scelto. La disponibilità di dati di questo tipo ha consentito non solo l'approfondimento di tematiche classiche delle discipline dei trasporti (i.e. scelta del percorso), ma anche l'apertura ad ambiti diventati di interesse rilevante solo recentemente. In particolare, i dati GPS sono diventati fondamentali per analizzare gli spostamenti effettuati con mezzi diversi dall'auto (a piedi, in bici etc.), per condurre indagini per periodi più lunghi di un giorno (*panel analysis*), nell'analisi delle emissioni ed infine per valutare l'efficacia di politiche mirate a promuovere la sostenibilità dei trasporti, in termini di riduzione dei chilometri percorsi prima e dopo un determinato intervento.

Proprio in quest'ultimo ambito, il CRiMM (Centro ricerche Modelli di mobilità) dell'Università di Cagliari, da alcuni anni sta conducendo sul territorio cagliaritano degli studi volti ad analizzare i comportamenti di viaggio dei cittadini con l'obiettivo di individuare soluzioni utili a ridurre l'uso dell'autovettura privata e favorire modalità di trasporto alternative. Durante una prima indagine pilota condotta tra

Luglio e Ottobre 2009, ad un gruppo di studenti dell'Università di Cagliari è stato fornito uno smart phone dotato di modulo GPS, sul quale è stata pre-installata un'applicazione, denominata "*Activity Locator*", per raccogliere in tempo reale dati di attività e viaggio. Questi due elementi – GPS e *Activity Locator* - hanno consentito di ottenere due tipologie di informazioni, già organizzate dal punto di vista informatico per essere successivamente manipolate: (1) informazioni spaziali attraverso le quali identificare il percorso seguito da ogni utente con i diversi modi, e le varie destinazione e (2) attributi dello spostamento sia qualitativi (scopo dello spostamento, modo utilizzato, compagnia, etc.) che quantitativi (tempo dedicato alle attività e agli spostamenti, distanze percorse).

La qualità dei dati GPS raccolti, in termini di coordinate esatte vs. errate o mancanti,- così come in qualsiasi indagine con dispositivo GPS - è risultata vincolata all'efficienza tecnologica del dispositivo utilizzato (errori nell'orologio dei satelliti, errori dovuti a disturbi sul ricevitore, tipo di ricevitore), e al contesto in cui si è svolta l'indagine (numero di satelliti disponibili, condizioni atmosferiche, ambiente costruito caratterizzato da palazzi troppo alti, ponti e tunnel). Questi aspetti, che in generale conducono a problemi di diversione delle coordinate rilevate dal modulo GPS rispetto al tratto stradale realmente impegnato, sono stati evidenziati attraverso la sovrapposizione dei percorsi rilevati su un grafo di rete esistente della città di Cagliari.

In questa tesi di dottorato, si riporta uno studio sistematico ed articolato nel campo della matematica, dell'informatica, delle telecomunicazioni, della statistica e della letteratura specifica nell'ambito dei Sistemi di Trasporto Intelligente (ITS) riguardante il problema dell'esatta proiezione e ricostruzione di un percorso reale su una mappa digitale. Negli ultimi anni la comunità scientifica si è interessata sempre

più e con risultati incoraggianti alla ricerca di soluzioni che potessero risolvere i problemi relativi all'errata localizzazione dei percorsi rilevati mediante rilevazioni GPS, ovvero del problema relativo all'implementazione della cosiddette procedure di *Map Matching* (MM).

Il MM utilizza le informazioni spaziali rilevate dal GPS e le integra con quelle di una qualsiasi mappa digitale: se entrambe le informazioni non fossero affette da errori (di rilevamento o di costruzione della mappa) la procedura per definire l'esatta collocazione del veicolo consisterebbe in una semplice proiezione della posizione del veicolo sul tratto di strada della rete. Nella pratica tuttavia gli errori di rilevazione dovuti al sistema GPS e quelli del sistema georeferenziato, nonché la complessità della topologia della rete stradale, le condizioni del traffico e le variazioni dei tempi di percorrenza, la presenza di tratti curvilinei o rettilinei generano procedure di MM che possono assumere connotazioni più o meno complesse a seconda del caso specifico di applicazione e a seconda delle informazioni disponibili per l'implementazione dell'algoritmo. Se ad esempio si conosce con certezza il punto di partenza, il percorso e il punto di arrivo di un autobus di linea, la procedura è estremamente semplice; ma se non si possiede nessuna delle informazioni precedenti, l'algoritmo assume delle connotazioni di complessità che esigono lo sviluppo di metodologie complesse.

L'evoluzione delle strategie di MM è avvenuta nel tempo in funzione del miglioramento delle tecnologie hardware che hanno permesso di migliorare la ricezione GPS, della possibilità di ricezione di ulteriori segnali di frequenza satellitari prima non disponibili per motivi di sicurezza nazionale, dello sviluppo informatico che tra l'altro ha permesso di analizzare una quantità di dati sempre maggiore in tempi computazionali prima irraggiungibili. Ma anche la maggiore richiesta di

quel complesso di servizi e di applicazioni rivolte all'aumento della sicurezza, alla riduzione del traffico, al miglioramento della mobilità, hanno creato nuovi scenari con cui confrontarsi e di cui soddisfare le richieste. La possibilità di poter analizzare dati in tempo reale ha accelerato lo sviluppo dei sistemi di trasporto telematici avanzati (ATTS) e conseguentemente la ricerca si è sempre più interessata alla risoluzione di problematiche specifiche: sistemi di controllo dinamico del traffico, di incidenti o di condizione della rete stradale sono solo alcuni esempi.

La tesi si sviluppa analizzando le motivazioni che in letteratura hanno spinto diversi ricercatori ad affrontare le problematiche che scaturiscono dalla necessità di disporre di informazioni corrette di tracciamento spaziale dei percorsi, ed in particolare introducendo le procedure di MM e le diverse teorie che si sono sviluppate intorno al problema, passando poi a proporre un contributo metodologico con un'applicazione informatica sviluppata con il linguaggio open source R.

La tesi si articola in quattro capitoli. Nel primo capitolo viene descritto il sistema di posizionamento globale, più comunemente conosciuto con l'acronimo GPS (Global Positioning System), diventato nelle ultime due decadi una tecnologia di uso comune. Nato come sistema satellitare di rilevazione americano per scopi militari negli anni della guerra fredda, ha visto la sua applicazione in campo civile agli inizi degli anni '90 e nel giro di pochi anni la sua applicazione è diventata così determinante che, dal rapporto del Federal Radionavigation Plan del 2000 a cura del Governo degli Stati Uniti (DoD) e dei Trasporti (DoT), se ne riconosce la crescente importanza civile e commerciale nello sviluppo economico del paese e dell'economia mondiale, nonché di importanza cruciale per molti settori infrastrutturali a livello globale.

La prima parte del capitolo descrive il GPS in relazione ai tre segmenti del sistema ed in particolare al sistema satellitare, i motivi che comportano una degradazione del segnale, gli errori che possono presentarsi e la loro valutazione.

L'altro componente importante è la mappa georeferenziata digitale in quanto le informazioni satellitari devono trovare corrispondenza su un reticolo stradale cartografico esistente: gli errori dovuti a mancanza di informazioni, al processo di digitalizzazione o agli errori di scala possono compromettere la corretta attribuzione del percorso stradale compiuto, alla reale rete stradale. La seconda parte del capitolo riguarda quindi i problemi relativi alla cartografia ed in particolare i diversi metodi di proiezione piana e gli inconvenienti che possono sorgere.

Nel secondo capitolo vengono analizzati i diversi approcci al MM, classificati secondo quanto proposto da diversi autori dal punto di vista metodologico e in relazione alla frequenza temporale della rilevazione del percorso, alle rilevazioni e l'analisi in tempo reale o differito, alla complessità del sistema stradale, ai diversi parametri che possono essere presi in considerazione. In particolare si sono descritti i diversi approcci delle metodologie del MM ed in particolare le tre tipologie di quello geometrico (point to point, point to curve e curve to curve), topologico e gli algoritmi avanzati. Si descrive, inoltre, il procedimento alla base dell'algoritmo topologico esponendo la proposta di Greenfeld (2002) e quella di Quddus (2003), il cui confronto permette di analizzare più nel dettaglio le diverse difficoltà che si possono incontrare nelle procedure di MM.

Definire una rigida catalogazione delle metodologie non permette di cogliere la complessità delle situazioni che possono essere analizzate, in

quanto entrano in gioco una molteplicità di circostanze. Oltre alla suddivisione rispetto al metodologia scientifica utilizzata, si è riportata anche quella che si basa sulla caratterizzazione rispetto alla quantità della rilevazione del posizionamento del veicolo mediante GPS in un arco di tempo, per la quale le procedure MM si possono suddividere in algoritmi a bassa o ad alta frequenza di rilevamento, rispetto alla tipicità della rete stradale, per cui si identificano algoritmi per reti stradali ad alta o bassa complessità; rispetto al tempo di analisi per la risoluzione dei problemi on-line o off-line.

Aldilà delle diverse suddivisioni possibili, ci sono dei punti critici che accomunano tutti i tipi di approcci: la necessità di individuare con precisione il primo punto rilevato tramite GPS alla mappa digitale. La maggiore complessità della soluzione in presenza di incroci, che diventa tanto più complessa quanto più è articolato l'incrocio stesso; la perdita di precisione delle informazioni GPS a basse velocità di viaggio; la diminuzione di precisione quando la rete stradale è molto complessa, come nel caso di centri cittadini altamente urbanizzati. Inoltre, la prestazione degli algoritmi proposti dipende dalla qualità della mappa digitale e dal ricevitore GPS: se un algoritmo è stato sperimentato in un centro urbano a bassa densità e con una mappa digitale con una certa scala di rappresentazione, non è detto che abbia lo stesso risultato applicato in un contesto diverso.

Il terzo capitolo è dedicato alla descrizione di metodologie più avanzate, ovvero all'approccio probabilistico, mediante la descrizione di un algoritmo funzionale per percorsi a bassa frequenza di rilevamento; la logica e il modello fuzzy è stata utilizzata da diversi autori tra cui Quddus et al. (2006), Syed e Cannon (2004) e Zhang e Gao (2008) di cui si confrontano gli algoritmi proposti. Infine l'ultima parte del capitolo

riguarda l'analisi di una proposta di risoluzione mediante l'utilizzo di algoritmi basati sulla densità. Poiché gli algoritmi avanzati fanno riferimento a concetti matematici e metodologie specifiche, nell'appendice si riprendono i concetti base del filtro Kalman, della logica e del modello fuzzy e del data clustering e la teoria dell'evidenza.

Nel quarto ed ultimo capitolo si espone una proposta metodologica al problema, che si basa sul confronto tra un percorso GPS nella sua totalità e un set di strade candidate che ricadono in una finestra definita intorno al percorso stesso. Grazie ai singoli punti rilevati, si definisce una spezzata che rappresenta l'andamento del veicolo; intorno a tale percorso si definisce un'area di osservazione o finestra, che sarà funzione anche di un parametro definito a priori corrispondente a quanto si vuole ampliare l'indagine. Questa finestra conterrà il percorso GPS e una serie di strade della mappa digitale che rappresenteranno i candidati al *map matching*: le aree tra il percorso e ogni singola strada vengono interpretate come un costo che si sostiene per ritornare su un tracciato stradale definito; quella strada che renderà minimo tale costo rappresenterà la strada migliore. La procedura è stata sviluppata in ambiente R, dapprima su dati sperimentali e successivamente su un percorso realmente rilevato durante l'indagine pilota del CRiMM di Cagliari.

La procedura sviluppata sembra aver dimostrato che, utilizzare come punto di riferimento aree di "penalizzazione" associate ad ogni singola strada può essere un fattore discriminante per individuare la migliore strada candidata. Essa è stata progettata in maniera parametrica per poter compiere in futuro ulteriori test che rendano robusta la proposta e verificabile in particolari circostanze (cavalcavie, rotonde complesse, incroci plurimi, ecc). L'ipotesi di lavoro sviluppata può essere intesa come

un primo passo per uno sviluppo più articolato e sistematico di ulteriori approfondimenti futuri.

Desidero qui ringraziare innanzitutto il Prof. Italo Meloni che ha seguito il mio percorso di ricerca e di studio, per il supporto che mi ha dato e la possibilità di collaborare con lui e il suo gruppo di lavoro al CRiMM. Non da meno è stato il suo supporto morale che ha reso questo percorso di studio e di ricerca realizzabile. Desidero altresì ringraziare l'ing. Erika Spissu per i consigli e l'aiuto fornitomi nello svolgimento per la ricerca sull'Activity Locator. Ringrazio inoltre l'Ing. Benedetta Sanjust di Teulada per il prezioso contributo datomi a seguito delle riflessioni comuni sulle problematiche di *map-mathing*.

Ringrazio infine la mia famiglia per la comprensione che ha mostrato nell'accettare che il tempo abitualmente loro dedicato, in special modo in quest'ultimo periodo, sia stato molto minore.

1. Il rilevamento della posizione e la mappa digitale

Il Sistema Informativo Geografico (GIS) si configura come un sistema di informazioni riferibili al territorio, quindi georeferenziato e insieme al sistema di posizionamento globale, più comunemente conosciuto con l'acronimo GPS (*Global Positioning System*), sono diventati strumenti di uso comune per compiere analisi e prendere decisioni operative. Il GPS, pur essendo il sistema di rilevamento maggiormente usato, è affetto però da errori sistematici e casuali, ovvero errori dovuti ai satelliti trasmettenti, alla propagazione dei segnali nell'etere, alle modalità della ricezione delle basi a terra del sistema o allo strumento ricevente. Inoltre, le caratteristiche del contesto fisico del territorio (tunnel, effetto canyon, etc.) possono determinare una perdita di qualità e di precisione dei dati. Spesso per attenuare alcuni di questi problemi si fa uso di ulteriori sensori riceventi che permettono di migliorare l'accuratezza delle informazioni, o di tecniche matematiche che tendono ad eliminare i "rumori" del segnale. L'altro componente importante è la mappa georeferenziata digitale, in quanto le informazioni satellitari devono trovare corrispondenza su un reticolo stradale cartografico esistente: gli errori dovuti a mancanza di informazioni, al processo di digitalizzazione o agli errori di scala possono compromettere la corretta attribuzione del percorso stradale compiuto, alla reale rete stradale. Il capitolo descrive il GPS in relazione ai tre segmenti del sistema ed in particolare al sistema satellitare, i motivi che comportano una degradazione del segnale, gli errori che possono presentarsi e la loro valutazione. La seconda parte del capitolo riguarda i problemi relativi alla

cartografia ed in particolare i diversi metodi di proiezione piana e gli inconvenienti che possono sorgere.

1.1 La tecnologia GPS

Il sistema di navigazione satellitare globale (GNSS) è l'insieme di satelliti che trasmettono segnali radio ad alta frequenza e determinano la posizione tridimensionale di un punto sulla Terra e intorno ad essa. I sistemi come il *Global Positioning System* sviluppato dal Dipartimento della Difesa americano, il *GLObal NAVigatsionnaya Sputnikovaya Sistyema* (GLONASS) russo, composta da 30 satelliti in costellazione ma solo 23 operativi e il sistema Galileo europeo, sono esempi di sistema di navigazione satellitare. Quest'ultimo nell'ottobre 2011 ha mandato in orbita i primi due satelliti e se ne prevedono 30 satelliti entro il 2014.

Già oggi è possibile utilizzare sia la copertura GPS che GLONASS e si presume che sarà possibile utilizzare in futuro anche il sistema GALILEO che quello americano in modo da usufruire di 60 satelliti e quindi offrire una migliore copertura nelle regioni europee e soprattutto quelle a più alte latitudini.

Le possibili applicazioni di tali sistemi sono rivolte ai rilevamenti terrestri (topografici, geodetici, posizionamento di veicoli in tempo reale, monitoraggio di deformazioni del suolo, ecc.), a quelli marittimi (studi sulle correnti marine, navigazione marittima, posizionamento di navi in tempo reale, ecc.), ai rilievi spaziali e aerei (posizionamento di veicoli in tempo reale, sistemi di navigazione, rilievi topografici, geometrici, rilevazioni di altri satelliti, ecc.).

Il sistema di posizionamento globale o *Global Positioning System* (GPS) è il sistema di radionavigazione satellitare Statunitense ad opera del Dipartimento della Difesa americana che fornisce 2 livelli di servizio GPS: il *Precise Positioning Service* (PPS) e lo *Standard Positioning Service* (SPS). Il primo è destinato esclusivamente ad utenti autorizzati dagli Stati Uniti e dalle strutture militari degli alleati americani le cui trasmissioni avvengono mediante un codice criptato P(Y); il secondo meno accurato del precedente, disponibile per la Comunità internazionale.

Le principali caratteristiche del sistema GPS SPS sono la copertura globale sino ad una altitudine di 3000 km e gratuita per tutti i suoi utenti; una precisione del 95% delle informazioni (di 13 metri sul piano orizzontale e 22 metri su quello verticale); la disponibilità continua nel tempo. Il sistema GPS è composto da tre segmenti: spaziale, terrestre o di controllo e utente.

L'attuale segmento spaziale è composto da 24 satelliti operativi e 3 latenti che vengono attivati in caso di guasto dei principali, ed orbitano a circa 20 chilometri dal suolo terrestre; sono disposti su 6 piani orbitali ognuno dei quali ne contiene 4 come mostra la Figura 1.

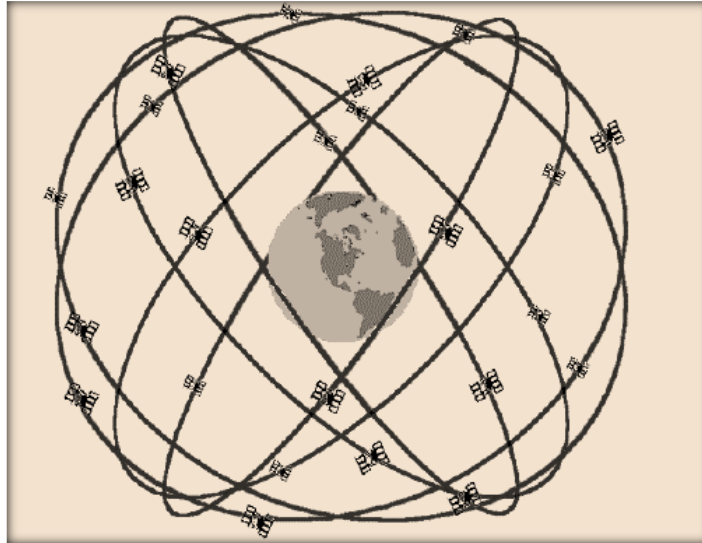


Figura 1: Piani orbitali

Il secondo segmento è il controllo terrestre composto da una stazione principale (detta *Master Station Control* (MSC), una di *back up* (BMCS), sei stazioni di monitoraggio (MS) e 4 antenne (Ground Antenna). La MSC ha la funzione principale di assicurare un riferimento comune mediante una serie di orologi atomici. Le cinque stazioni di controllo ricevono continuamente i segnali emessi dai satelliti, verificano i dati descrittivi degli stessi (effemeridi) e predicono l'orbita di ciascun satellite in un determinato intervallo. Le informazioni così aggiornate vengono trasmesse ai satelliti tramite le tre antenne di terra, controllate a distanza.

Il segmento utente o di utilizzo è costituito dagli utenti del servizio GPS; un ricevitore demodula i segnali emessi dai satelliti per stimare la propria posizione in tempo reale. I demodulatori più utilizzati decodificano le portanti L1 e L2 del segnale GPS, di cui si parlerà successivamente.

1.2 Il segmento spaziale

I satelliti non sono geostazionari in quanto percorrono l'orbita completa intorno alla terra in 12 ore; il controllo da terra è garantito da 5 stazioni posizionate sul territorio americano che provvedono al sincronismo dei diversi parametri temporali e di navigazione dei satelliti.

Il segnale che parte dai satelliti è una combinazione di tre componenti: due segnali multipli del segnale fondamentale $f_0 = 10.23 \text{ MHz}$, i codici C/A (*Coarse/Acquisition*) e il codice P (*Precision*) e il terzo componente detto Messaggio Navigazionale (NM). I due segnali $f_1 = 154 \cdot f_0$ e $f_2 = 120 \cdot f_0$ hanno una lunghezza d'onda pari rispettivamente a circa 19 cm e 24 cm. I segnali GPS sono fortemente direzionali e sono facilmente riflessi o bloccati da oggetti solidi o da superfici di acqua. I codici C/A sono sequenze di 1023 bit, (*noise pseudo-casuali*) con lunghezza d'onda di circa 300 metri non sono criptati e quindi sono disponibili per l'accesso civile e trasmessi solo sulla frequenza L1; ogni satellite possiede un differente codice che viene ripetuto ogni millisecondo. Questo codice viene utilizzato dagli utenti militare per acquisire il codice P. Il codice P ha una frequenza di 10.23 MHz, anch'esso definito come noise pseudo casuale, con una lunghezza d'onda di circa 30 metri, e un periodo di 267 giorni. Viene trasmesso sulle frequenze d'onda L1 ed L2; viene utilizzato esclusivamente per uso militare e richiede un ricevitore particolare per demodulare il segnale.

Anche il *Navigation Message* è trasmesso sulle frequenze L1 e L2, e contiene gli elementi orbitali del satellite ovvero il proprio

codice identificativo, i sincronismi temporali, i parametri di correzione del ritardo di propagazione ionosferica, i parametri di navigazione curati dal segmento di controllo da terra e l'Almanacco ossia le effemeridi di tutti gli altri satelliti. Il messaggio detto *frame* è suddiviso in cinque blocchi, detti *sub frame*; i dati dell'almanacco sono contenuti in un *sub frame* e consentono di determinare la posizione di tutti i satelliti e di conoscere quelli visibili e utilizzabili. L'almanacco è estremamente importante in quanto permette al GPS di selezionare quei satelliti più adatti in base alla loro posizione e per poter individuare con la massima precisione la posizione. I ricevitori GPS mantengono in memoria l'almanacco anche quando sono spenti e nel momento dell'accensione successiva hanno già a disposizione le informazioni per sincronizzarsi velocemente con i segnali dei satelliti desiderati. Nel caso in cui il periodo di spegnimento fosse notevole, o se i dati venissero persi, il ricevitore impiegherà più tempo ad individuare la sua posizione, in quanto dovrà ricevere nuovamente l'almanacco.

I livelli di utilizzo sono lo *Standard Positioning Service* (SPS) che utilizza la portante L1 ed è caratterizzato da una accuratezza di 100 metri sul piano orizzontale, e il *Precise Positioning Service* (PPS) che utilizza sia la portante L1 che L2 con una accuratezza di 18 metri sul piano orizzontale. Questa differenza di precisione ha avuto lo scopo di proteggere gli utenti privilegiati (i corpi militari) da trasmissioni di disturbo ed è dovuta alla degradazione del segnale da parte del Dipartimento della Difesa statunitense, che può produrre un errore fino a 100 metri senza alcuna informazione preventiva.

Per poter alterare la precisione del segnale sono utilizzate due tecniche: La *Selective Availability* (SA) che varia le effemeridi dei satelliti dalla stazione di controllo terrestre; nel 2000 è stato rimosso

tale disturbo. *L'Anti-Spooling (AS)* che agisce sul codice P ed introduce un codice criptato detto codice Y, di cui solo utenti privilegiati o autorizzati possono servirsene e decodificarlo con particolari ricevitori e programmi di decifrazione.

Nel maggio del 2000 gli Stati Uniti riconoscono la crescente importanza civile e commerciale del GPS nella crescita economica e nello sviluppo economico, come mezzo per migliorare la vita sociale e come il sistema sia di importanza cruciale per molti settori di infrastruttura. Per tale motivo nel 2007 il governo americano decide di promuovere una nuova generazione di GPS senza la selettività del segnale alla sola frequenza L1 e assicurare una continuità e globalità del servizio ad uso civile, visto anche la crescente dipendenza di tutti i servizi economici al GPS. Per modernizzare il sistema in aggiunta al segnale L1 (C/A) il USG ha introdotto altri tre tipi di segnale: L1C, L2C e L5 per il miglioramento delle applicazioni in campo civile. In campo militare verrà utilizzato un segnale di nuova generazione detto M-code, separato dal settore civile e da altri segnali satellitari di radiocomunicazione. L'obiettivo della modernizzazione è di migliorare l'accuratezza sia del posizionamento che della sincronizzazione, la globalità del segnale e la sua integrità e non ultimo i sistemi di controllo. I ricevitori a terra potranno essere ancora usati, ma per poter sfruttare al meglio le caratteristiche della nuova generazione di satelliti dovranno essere aggiornati i ricevitori.

Il segmento spaziale è caratterizzato da diverse fasi di sviluppo, detti blocchi: nella Tabella 1 sono evidenziati i diversi nomi dei

blocchi e il periodo del lancio in orbita, il numero dei satelliti del blocco e quelli attualmente in orbita e funzionanti.

I satelliti del primo blocco sono serviti ad eseguire i test necessari per il sistema; la loro vita media è risultata al di sopra delle aspettative (circa 7 anni rispetto alla vita media prevista di circa 4). Attualmente nessuno è funzionante; quelli del blocco II, IIA (*Advanced*) e IIR (*Replenishment*) degli anni '90 hanno presentato molti miglioramenti rispetto ai precedenti, tra cui una maggiore resistenza degli apparati elettronici alle radiazioni, una capacità di immagazzinamento dei dati di navigazione per 180 giorni, contro i circa 4 giorni dei satelliti del blocco I, la capacità di ricevere messaggi di aggiornamento da terra ogni 14 giorni e ad avere comunque una precisione accettabile, la capacità di definire le distanze tra un satellite e l'altro, il maggiore carico del propellente.

I satelliti IIR-M (*Modernized*) del nuovo secolo sono forniti di emettitori sperimentali che trasmettono su una nuova frequenza, L2C che permette di migliorare le applicazioni commerciali in campo civile e due ulteriori segnali, L1M e L2M con un nuovo codice (*M-code*) in campo militare. Quelli del blocco IIF (*Follow-on*) effettivamente lanciati in orbita ed operativi sono due e trasmettono un ulteriore segnale civile L5 per la sicurezza nei trasporti. Ogni satellite è fornito di un orologio atomico che consente una maggiore accuratezza nella navigazione.

Tabella 1: le diverse fasi di sviluppo e le caratteristiche

Blocco	Lancio	Numero Satelliti	Attualmente operativi
I	1978/85	11	0
II	1989/90	9	
IIA	1990/97	19	10
IIR	1997/2004	13	12
IIR-M	2005/09	8	7
IIF	2010/11	12	2(effettivamente lanciati)
IIIA	2014/...		

1.3 Gli errori di misura del GPS

L'errore di misura del GPS dipende oltre che dai sistemi di degradazione, anche da altri tipi di errori come quelli dovuti ad una bassa qualità tecnica del ricevitore, al ritardo del segnale dovuto alla cattiva propagazione negli strati atmosferici della ionosfera e troposfera, agli errori nella determinazione delle orbite e ai parametri di correzione degli orologi satellitari introdotti dal sistema di controllo terrestre. Il ritardo atmosferico può comportare un errore di posizionamento che può variare dai 5 ai 200 metri. I disturbi possono essere dovuti al percorso di curvatura ottica (*optical path curvature*) in quanto ogni segnale elettromagnetico segue la legge di Fermat secondo cui il segnale segue il percorso minimo che non coincide necessariamente con la distanza minima geometrica.

Gli errori dovuti alla rifrazione della troposfera dipendono dai parametri dell'atmosfera (temperatura, vapore acqueo, pressione, ecc.) e dall'angolo zenitale del satellite. Questi due tipi di errore non possono essere eliminati e quindi sono uguali per entrambe le frequenze L1 ed L2. L'errore aumenta se l'angolo zenitale supera i 75°, così come aumenta se il clima è caldo e umido. I ritardi ionosferici sono dovuti alla densità degli elettroni liberi contenuti nella parte alta dell'atmosfera, ad una altezza oltre i 40 km; il disturbo dipende dal tipo di segnale e quindi è diverso per le due portanti. Il disturbo si presenta più o meno intenso a seconda se il periodo considerato è notturno o diurno o in relazione all'attività solare.

Un ulteriore errore è quello prodotto dalla combinazione di segnali diretti e riflessi: questi prendono il nome di errori multipli (*multipath*). Poiché dipendono dal luogo di localizzazione e dalla presenza o meno di grosse superfici di acqua, non è possibile prevederne l'ampiezza ma sicuramente è possibile ridurli o eliminarli. Con l'effetto multipath l'errore può superare anche i 50 metri. Nella Tabella 2 vengono riportati gli errori quadratici medi residui degli errori che permangono dopo le possibili compensazioni e correzioni degli errori eliminabili.

L'errore quadratico medio complessivo è di circa 33 metri per i ricevitori degli utenti non privilegiati, mentre è di circa 6 metri per quelli che decodificano il codice P. Ma l'errore di posizionamento dipende anche dalla configurazione geometrica dei satelliti; per valutare la bontà di tale configurazione sono definiti alcuni indici che definiscono la degradazione della precisione. Valori bassi rappresentano una buona configurazione dei satelliti. L'indice di

degradazione della precisione nella determinazione tridimensionale del punto è detta *Position Dilution Of Precision* (PDOP). Questo è un fattore di scala ed è legato alla distribuzione in azimut dei satelliti e alla loro distribuzione in elevazione. Il PDOP può assumere un valore minimo pari a 1,6: se il valore è basso la distribuzione in azimut dei satelliti utilizzati per le misure di distanza è ottimale; se assume valori tra 2 e 4 la distribuzione è normale altrimenti i satelliti utilizzati sono molto concentrati e quindi l'errore risulta essere più alto. la Figura 2 illustra un esempio di diversa configurazione geometrica dei satelliti.

Tabella 2: I tipi di errore e l'errore associato

Tipi di errore	Errore quadratico medio residuo
Orologi satellitari	3
Effemeridi	3
SA	30
Ritardi ionosferici	9
Ritardi troposferici	2
Rumore ricevitore	10
Multipath	3

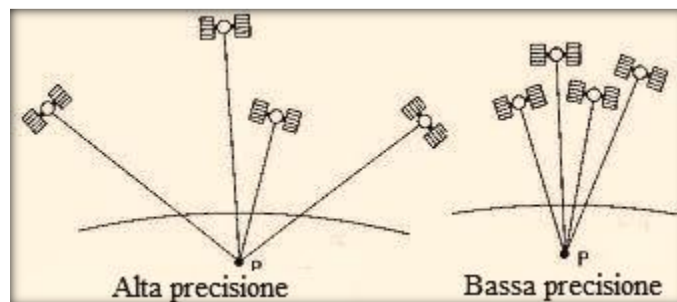


Figura 2: Esempi di configurazioni satellitari

Se si potesse decodificare sia le trasmissioni sulla gamma radio L1 che sulla gamma radio L2 il margine di errore varierebbe tra 1 e 0.5 metri. Ciò non è tuttora possibile in quanto un ricevitore GPS per uso civile riceve solo segnali sulla frequenza L1, in quanto l'altra è preclusa all'utenza civile: questa restrizione porta l'errore di rilevazione a circa 15 metri. Se il ricevitore rileva anche l'altitudine sul livello del mare il margine di errore aumenta a 20-25 metri.

1.4 Misure di distanza

Il codice *Coarse Acquisition* (C/A) si ripete ogni millisecondo ma viene captato con un ritardo t , pari al rapporto tra la distanza tra il satellite e il ricevitore e la velocità della propagazione delle onde elettromagnetiche. Il ricevitore possiede un generatore locale del codice C/A che dovrebbe essere sincronizzato con il riferimento temporale (*GPS Time*); questo sincronismo però non è perfetto in quanto gli orologi utilizzati nei ricevitori sono generalmente dotati di oscillatori al quarzo. Le misure di distanza operate dai ricevitori sono note come misure di *pseudorange*. I segnali provenienti dal satellite e dal ricevitore sono confrontati mediante autocorrelazione: il dispositivo che opera il confronto è detto correlatore e produce un segnale funzione dello sfasamento dei due segnali: se il valore è massimo i due segnali sono perfettamente allineati, mentre se il valore è nullo sono disallineati. Per poter ottenere un perfetto sincronismo il correlatore genera un ritardo in modo da ottenere sempre un valore massimo. La misura del ritardo permette di calcolare il tempo di viaggio del segnale e quindi la distanza del satellite. Per esempio, con un errore temporale di $0,1 \mu\text{sec}$ l'errore di misura della distanza sarà pari a 30 metri. Il costo dei ricevitori

dotati di orologi di precisione è notevole, ma la precisione può essere recuperata determinando l'errore rispetto al segnale del tempo proveniente dal satellite.

Non basta conoscere la distanza di un punto di rilevazione ad un solo satellite, in quanto la distanza dal primo satellite individua la posizione del ricevitore sulla superficie di una sfera centrata sul satellite. Neanche la conoscenza della distanza tra due satelliti è sufficiente in quanto la determinazione della seconda distanza permette di individuare il punto di intersezione delle due sfere. In definitiva è necessario conoscere la distanza da almeno tre satelliti, in quanto l'intersezione di tre sfere determina due punti, uno dei quali è ad una quota molto alta e quindi da scartare. Con la misura della distanza ad un altro satellite il punto è determinato con certezza. Questo procedimento è conosciuto come trilaterazione (Figura 3).

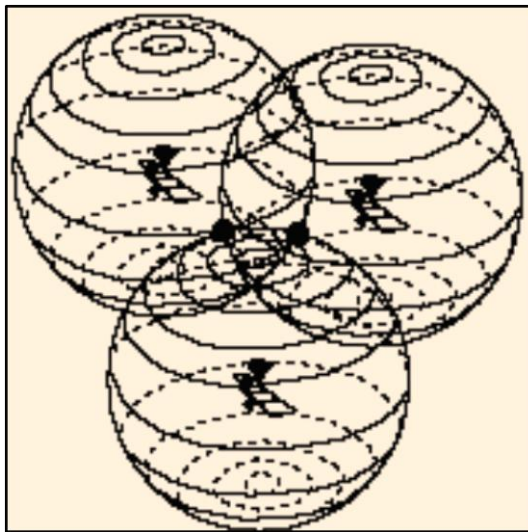


Figura 3: Trilaterazione

1.5 La mappa digitale

Di fondamentale importanza rivestono le informazioni relative al territorio su cui si vuole rilevare uno spostamento spaziale e temporale. Infatti le informazioni che vengono acquisite con la rilevazione satellitare devono essere confrontate con quelle che definiscono la topologia e la composizione stradale del percorso di un generico viaggio. Errori della mappa producono ulteriori errori in tutto il processo di rilevamento di un percorso mediante un dispositivo GPS; è necessario quindi rappresentare correttamente la scala, il livello di generalizzazione, la proiezione, il datum e le coordinate del sistema.

La scala può essere definita come il rapporto tra la misura effettuata sulla mappa e la corrispondente misura reale ed è indicata con $1:M$ in cui M è la misura reale rispetto al tratto unitario. Generalmente si identificano carte di riferimento in piccola scala (più piccola di 1:100.000) in cui si rappresenta un vasto territorio con pochi particolari, in media scala (1:24.000 fino a 100.000) e in larga scala (1:1 fino a 1:24.000) che rappresenta una piccola parte di territorio con molti particolari. L'accuratezza di particolari è necessaria quando l'indagine si riferisce ad aree urbane complesse, come ad esempio città o metropoli con un reticolo stradale molto articolato e complesso; meno indispensabile quando il territorio non possiede una struttura articolata, come le aree rurali o a bassa densità abitativa.

Per motivi di sinteticità non si riescono ad identificare le strade o le indicazioni relative ad incroci o rotatorie se non con un tratto lineare, più o meno curvato, che non permette l'identificazione

dell'ampiezza delle strade: lo spessore minimo del tratto grafico viene assunto per convenzione pari 0,2 millimetri, e se si usa una scala 1:10.000 si avrà una corrispondenza reale pari a due metri, che rappresenta la lunghezza minima del particolare; l'errore massimo è generalmente di 0,5 millimetri che corrisponde a 5 metri. Esiste quindi una incertezza nel posizionamento di un punto su una mappa corrispondente a 0,5 millimetri. Se poi si vuole passare da una mappa 1:5.000, per la quale l'errore di posizionamento è di 2,5 metri, ad una con una scala di 1:10.000, l'errore è inferiore a quello della mappa stessa se fosse prodotta in versione originale. Ciò non avviene se vogliamo ulteriormente ingrandire il particolare: se si passa ad una scala 1:2.000 l'errore rimane quello della scala di partenza di 2,5 metri mentre diversamente sarebbe stato di 1 metro.

1.6 Tecniche di proiezione

La tecnica della proiezione pone in corrispondenza biunivoca un sistema di coordinate di una superficie curva con un sistema di coordinate piane. Ma tutte le proiezioni sono affette da distorsioni: per effetto della deformazione e facendo riferimento ai meridiani e ai paralleli si possono commettere diversi tipi di errori relativi alla distanza (quando non esiste una corrispondenza alla scala dei rapporti tra le distanze sulla carta e quelle reali), alla superficie (se le estensioni uguali sulla carta non sono equivalenti ai territori reali) e agli angoli (se nella trasformazione non si conservano le ampiezze degli angoli). Rispetto allo scopo della rappresentazione si può pretendere per il mantenimento dei rapporti angolari, per cui i meridiani e i paralleli sono perpendicolari tra loro (proiezione conforme o isogona) o per la costanza dei rapporti tra le aree

(proiezione cosiddetta equivalente). Le proiezioni sono delle trasformazioni che riducono al minimo gli errori descritti in precedenza; in particolare si possono ricordare le proiezioni prospettiche e di sviluppo. Le prime sono proiezioni dirette che utilizzano metodi geometrici di trasposizione su un piano a due dimensioni e sono per questo dette proiezioni vere, mentre le seconde utilizzano una proiezione su un cono o un cilindro e successivamente su un piano. Nella Figura 4 sono rappresentate rispettivamente le proiezioni prospettiche vere polare, equatoriale ed obliqua, a seconda che il piano sia tangente al polo, all'equatore o ad un qualunque parallelo. La proiezione obliqua, produce in corrispondenza dei poli paralleli riconducibili a circonferenze.



Figura 4: Proiezioni prospettiche vere

Rispetto al punto di vista si possono ottenere diverse proiezioni: centrografica o gnomonica se il punto di vista è dal centro della Terra e i paralleli si allontanano molto dal centro; stereografica se il punto di vista è sulla superficie terrestre, l'ampiezza degli angoli si conserva nella proiezione, i paralleli si allontanano di poco dal centro e l'errore è meno accentuato; infine ortografica se il punto di vista è all'infinito e i raggi di proiezione sono paralleli.

Nella Figura 5 sono evidenziate le prospettive vere polari equatoriali rispetto ai tre punti di vista.

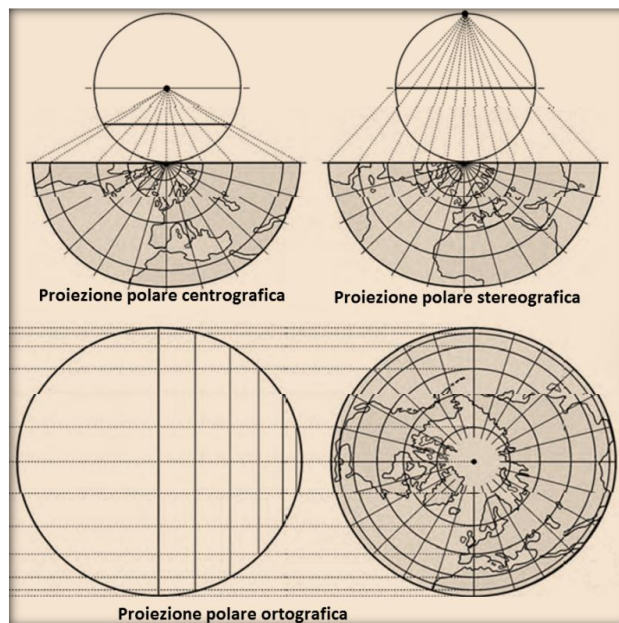
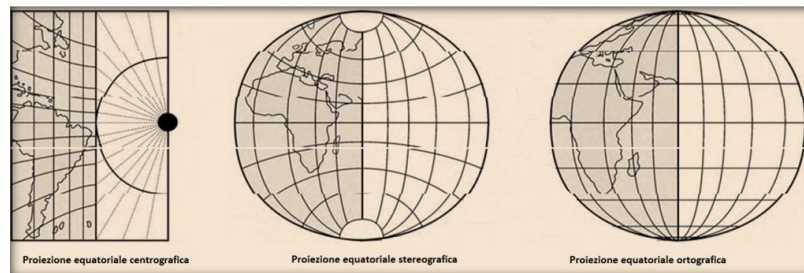


Figura 5: Proiezioni equatoriali e polari

Le proiezioni indirette producono trasformazioni su una superficie piana mediante proiezioni prospettiche su superfici sviluppabile su un piano, ovvero utilizzando un cilindro o un cono e successivamente proiettandole su un piano. Come si osserva dalla Figura 6 le proiezioni coniche sono create disponendo un cono intorno alla superficie terrestre; se si dispone un cilindro verticale

si produrrà una proiezione cilindrica diretta, mentre se il cilindro è orizzontale si otterrà una proiezione cilindrica inversa.

Nelle proiezioni cilindriche centrale diretta la superficie sferica è proiettata su un cilindro ad esso tangente e il centro di proiezione è il centro della sfera; l'asse del cilindro coincide con l'asse Nord-Sud.

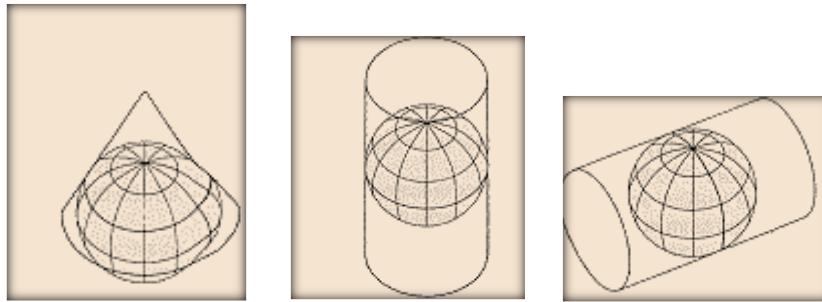


Figura 6: Proiezione conica, cilindrica diretta ed inversa

La proiezione di Mercatore si può considerare come una sua derivazione, in cui meridiani sono rappresentati da segmenti equidistanti a parità di differenza di longitudine, paralleli tra loro e perpendicolari all'equatore e ai paralleli. La loro distanza è corretta per ridurre al minimo gli errori angolari; in questa trasformazione si conserva il rapporto tra le aree della superficie terrestre e le aree della loro rappresentazione, ma la superficie si deforma sempre più in prossimità dei poli: per questo è applicata alle medie latitudini.

La proiezione di Gauss deriva da quella di Mercatore ed è una proiezione cilindrica inversa, anche chiamata Proiezione Inversa di Mercatore, che utilizza un reticolo cilindrico con asse perpendicolare all'asse terrestre. Il sistema *Universal Transverse Mercator* (U.T.M.) è la proiezione di Gauss utilizzata nella cartografia

mondiale; in particolare questo sistema di riferimento divide il globo in 60 fusi ognuno di 6° di ampiezza di longitudine e 160° di latitudine e usa un cilindro tangente al meridiano centrale del fuso; ogni fuso è identificato da una lettera dell'alfabeto.

Il parametro di trasformazione, ovvero il *datum*, è definito dall'ellissoide di riferimento e dal suo orientamento rispetto alla superficie reale. Gli ellipsoidi sono modelli del Globo terrestre sviluppati nel corso del tempo e migliorati rispetto al progresso dei sistemi di rilevamento; ogni ellissoide è identificato da un nome e dall'anno di ideazione. Per migliorare l'accuratezza si utilizzano sia *datum* locali che globali. I primi migliorano l'accuratezza di una particolare zona e si riferiscono ad un orientamento in un punto particolare, mentre per i secondi gli ellipsoidi sono orientati al centro della Terra. Fino agli anni '80 in Italia è stato utilizzato il *datum* locale Roma40 con l'ellissoide *Hayford* (Internazionale 1924) orientato a Roma Monte Mario (con azimut sul Monte Soratte) e successivamente il *datum* locale *European Datum 1950* (ED50), sempre con lo stesso ellissoide ma orientato in Germania a Postdam.

1.7 La proiezione di Gauss-Boaga

In Italia si è utilizzato la proiezione di Gauss-Boaga, derivata da quella di Gauss dal geodeta Boaga: il territorio è compreso tra i meridiani 6° di longitudine Est e 18° di longitudine Est rispetto a Greenwich. Se consideriamo i due meridiani di longitudine 9° e 15° rispetto a Greenwich questi saranno i meridiani centrali di due fusi: il fuso Ovest (tra il meridiano 6° e 12° di longitudine) e il fuso Est

(tra il 12° e il 18° di longitudine). I sistemi di coordinate cartesiane piane corrispondenti ai fusi Est ed Ovest costituiscono il sistema Gauss-Boaga; essi hanno come asse delle ascisse l'equatore e come asse delle ordinate la trasformata nel piano del meridiano centrale del fuso Est e del fuso Ovest. Ma i punti che ricadono sull'asse delle ordinate dovrebbero avere coordinata pari a 0, così come i punti a sinistra dell'asse delle ordinate dovrebbero avere coordinata negativa.

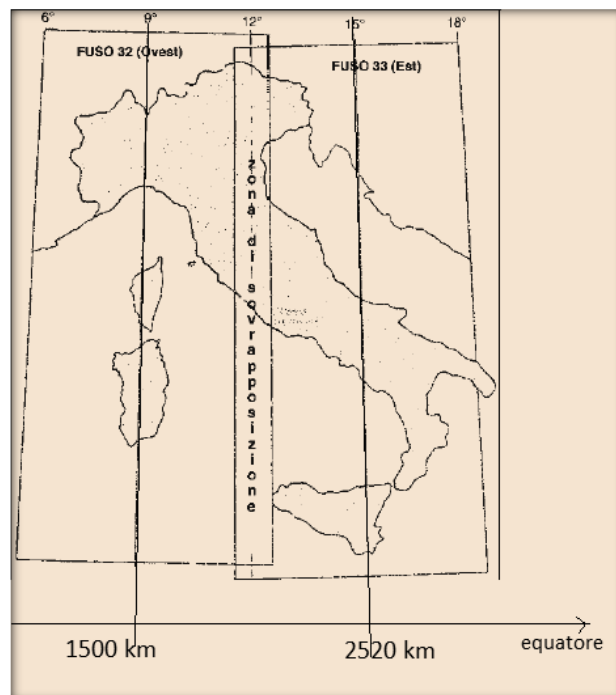


Figura 7: La proiezione di Gauss-Boaga dell'Italia

Per ovviare a questo inconveniente si somma un valore costante di 1500 Km alle coordinate del fuso Ovest e 2520 a quelle del fuso Est che sono identificate anche come false origini. Così i due meridiani origine hanno rispettivamente coordinate 1500 e 2529 Km. La prima cifra della coordinata permette di individuare subito il

fuso di appartenenza del punto in considerazione. Come si può notare dalla Figura 7 i due sistemi di coordinate presentano una zona sovrapposta in quanto il primo si spinge poco oltre i 12° mentre il secondo parte precisamente a 12° e termina oltre i 18° (18°30') per comprendere anche la penisola salentina.

Negli anni '50 si decise di adottare un sistema globale, ovvero l'U.T.M. In questa proiezione cilindrica inversa, come il sistema di Gauss-Boaga, la falsa origine è pari a 500 Km, i fusi sono numerati dall'antimeridiano di Greenwich e l'Italia è compresa tra i fusi 32, 33, e 34. Per l'emisfero Nord l'origine per la latitudine è pari a 0, mentre per l'emisfero Sud è 10.000.000; la distorsione aumenta da Est verso Ovest. Ogni fuso è composto da 20 fasce di 8° di ampiezza e l'intersezione tra la fascia e il fuso determina una zona costituita da quadrati di 100 Km di lato con rette parallele all'equatore e al meridiano centrale del fuso ; ogni zona è contraddistinta dal numero del fuso e dalla lettera indicante la fascia. In figura è evidenziata la penisola italiana identificata con il sistema UTM: essa si colloca nei fusi 32 S e T Nord, 33 S e T Nord.

Per contenere l'inevitabile deformazione nella percentuale del 4 per mille viene adottato per ogni fuso una correzione detta fattore di scala che in Italia è pari a 0.9996, ovvero si considera il meridiano centrale non nella sua vera lunghezza ma contratto di 0.996. Nella tabella 3 vengono riassunte le principali caratteristiche dei tre sistemi di riferimento utilizzati in Italia.

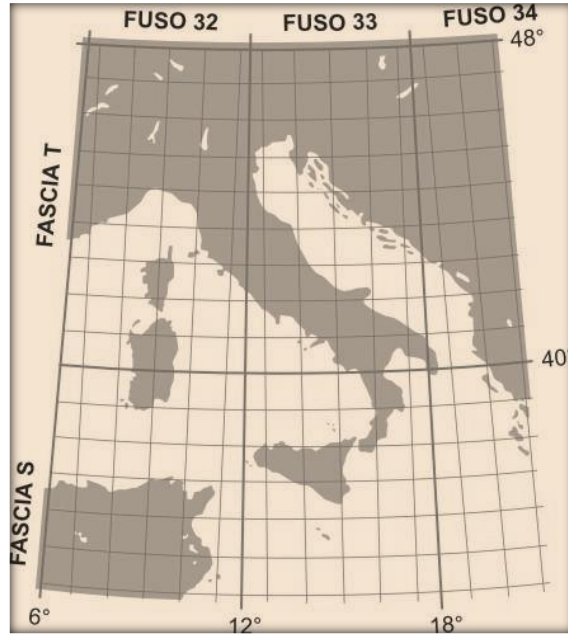


Figura 8: Proiezione U.T.M dell'Italia

Tabella3: I principali sistemi di riferimento italiani

<i>Datum</i>	<i>Ellissoide</i>	<i>Orientamento</i>	<i>Riferimento longitudinale</i>	<i>Riferimento latitudinale</i>	<i>Fusi</i>	<i>Meridiani centrali</i>	<i>False origini</i>	<i>Fattore di correzione</i>
Roma40	Hayford	Monte Mario	Monte Mario	Equatore	Est/Ovest	9° e 15° Greenwich	1500Km e 2520 Km	0.996
ED50	Hayford	Medio europeo	Greenwich	Equatore	32, 33, 34	9°, 15°, 21° Greenwich	500 km	0.996
WGS84	WGS84	Geocentrico	Greenwich	Equatore	32, 33, 34	9°, 15°, 21° Greenwich	500 km	0.996

1.8 Errori nella creazione della mappa digitale

Oltre agli errori dovuti alla creazione della mappa si possono presentare errori dovuti all'associazione degli attributi della rete nel passaggio delle informazioni al digitale e quindi alla formazione della mappa digitale. Una mappa digitale per essere definita di qualità deve possedere informazioni circa le rotatorie, gli incroci, le curve, ovvero le informazioni topologiche, ma anche la precisione di alcune informazioni come ad esempio la capacità di una rotatoria, i

suoi elementi geometrici o la loro classificazione, come ad esempio se è una rotatoria o un crocevia. Molte mappe digitali non forniscono informazioni relative ad esempio al numero di corsie, alla loro larghezza, al senso di marcia, informazioni che normalmente sono presenti su una mappa non digitale. La mancanza di questi specifici particolari può comportare errori di tipo geometrico che insieme agli errori topologici possono introdurre errori di posizionamento e di navigazione. Sicuramente con i software oggi a disposizione gli errori geometrici possono essere ridotti se non eliminati, mentre quelli di tipo topologico non sono così facili da correggere. La correttezza può essere misurata da indicatori statistici come la deviazione standard degli errori. Se si definiscono e_1, e_2, \dots, e_n le differenze tra le n misure della mappa e il loro vero valore, la correttezza della mappa può essere misurata dalla sua varianza, ovvero dalla media del rapporto tra il quadrato della somma delle differenze delle e_i il loro valor medio:

$$\sigma_M^2 = \frac{\sum_{i=1}^n (e_i - \mu)^2}{n} \quad \text{con} \quad \mu = \frac{\sum_{i=1}^n e_i}{n}$$

Sono stati proposti anche dei metodi empirici per misurare la precisione della mappa digitale. Ad esempio il Dipartimento di Geografia dell'Università del Texas (UTADG) ha calcolato l'accuratezza mediante la radice quadrata della somma dei quadrati di ogni componente. Nella tabella seguente è riportato un esempio di tale misura empirica, ipotizzando di stimare l'errore per una mappa in scala 1:2500. Al variare della risoluzione della mappa varierà in proporzione l'indicatore. Se, allora, si riporta uno stesso percorso su

due mappe digitali della stessa zona ma di diversa origine e in scala diversa, sarà molto probabile che il percorso sembrerà diverso

Tabella 4: Errore stimato in base alle componenti

Componente	Errore stimato in metri	c_i	$\sqrt{\sum c_i} =$ $= \sqrt{8.0625} = 2.839 \text{ metri}$
Documento di origine	2.5	6.25	
Importazione della mappa	1.25	1.5625	
Digitalizzazione	0.5	0.25	
Totale	8.0625		

2. Gli algoritmi di *map matching*

In questi ultimi anni si è assistito ad una forte richiesta e ad un crescente utilizzo dei Sistemi di Trasporto Intelligenti (ITS - *Intelligent Transport System*) come strumento chiave per l'attuazione delle politiche di trasporto. La localizzazione e il monitoraggio degli spostamenti sia per la gestione di flotte pubbliche che di veicoli privati avviene mediante dispositivi GPS a bordo dei veicoli, che consentono di acquisire informazioni spaziali dei movimenti in tempo reale al fine di compiere delle scelte operative o delle indagini esplorative. Essenziale risulta essere l'esatta attribuzione della posizione del veicolo rispetto alla mappa digitale, problema che in letteratura è conosciuto come *map matching* (MM). Il capitolo si apre con una panoramica del problema e dell'organizzazione delle informazioni dal punto di vista informatico; successivamente si ripercorrono gli sviluppi delle diverse metodologie di MM, differenziate per approccio geometrico, topologico e per l'utilizzo di metodi avanzati.

2.1 L'identificazione del percorso

Sin dalle prime applicazioni del GPS (Wagner, 1997) il passo successivo alla raccolta dati (*postprocessing*) è il cosiddetto *map matching* (MM), cioè l'associazione del percorso rilevato ad una mappa digitale, in genere al fine di identificare il tratto stradale in cui si trova il veicolo; si tratta quindi di identificare il percorso corretto che il veicolo percorre e determinare la sua posizione sul *network* di riferimento. Nelle figure

seguenti viene evidenziato un percorso seguito da un autoveicolo durante un test compiuto nella Città di Cagliari a cura del CIREM. I punti posizione (pp) GPS sono stati riportati sulla mappa stradale digitale della città di Cagliari ed resi visibili attraverso un visualizzatore *open source* denominato TatukGis (Figura 9); riportando poi i pp sulla mappa stradale di Google si può notare come il percorso GPS sia quasi sempre deviato rispetto dal tratto stradale (Figura 10).

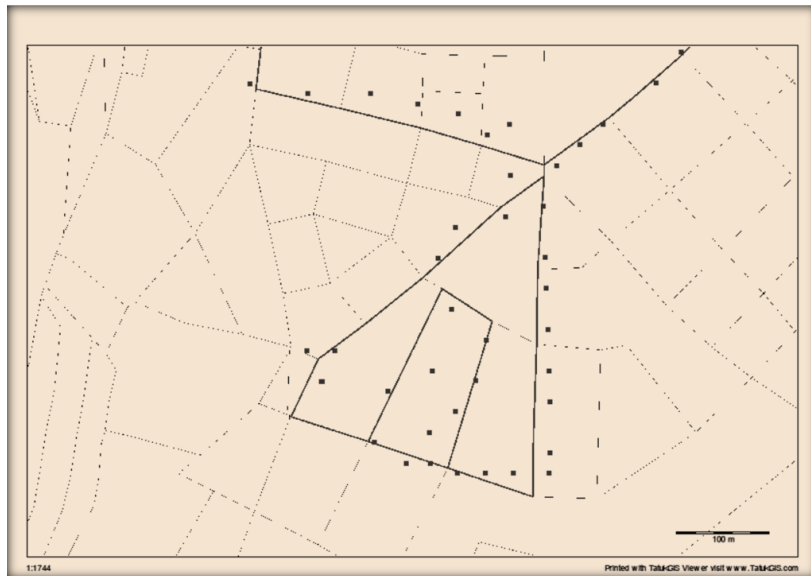


Figura 9: Percorso GPS e dopo il *map matching* con TatukGis



Figura 10: Percorso da tracciato GPS con mappa Google

I singoli punti sono le rilevazioni satellitari della localizzazione del veicolo nel tempo. Maggiore sarà l'intervallo tra una rilevazione ed un'altra minori saranno i punti rilevati. Un ricevitore GPS fornisce quindi informazioni più dettagliate se l'intervallo di tempo tra un segnale ed il successivo è piccolo. I punti rilevati P_i rappresentano i nodi di una ipotetica strada: congiungendo due nodi consecutivi A e B si ottiene un segmento lineare o arco: il percorso ricostruito è la creazione di una serie di archi, ovvero di polilinee.

Per quanto riguarda la mappa digitale le informazioni geometriche sono riferite alla linea centrale della strada (ovvero lo *shape*); quindi ogni elemento è sempre riferito all'informazione centrale della strada. I dati di base per la creazione di una mappa digitale sono le informazioni relative alle coordinate dei nodi, ovvero la latitudine e la longitudine, e le

informazioni relative agli archi, ovvero il numero identificativo dell'arco, il nodo iniziale e il nodo finale. Nelle Tabella 5e 6 vengono riportati parti dei file contenenti i nodi e gli archi della mappa digitale di Cagliari.

Tabella 5: Coordinate di alcuni nodi della rete stradale

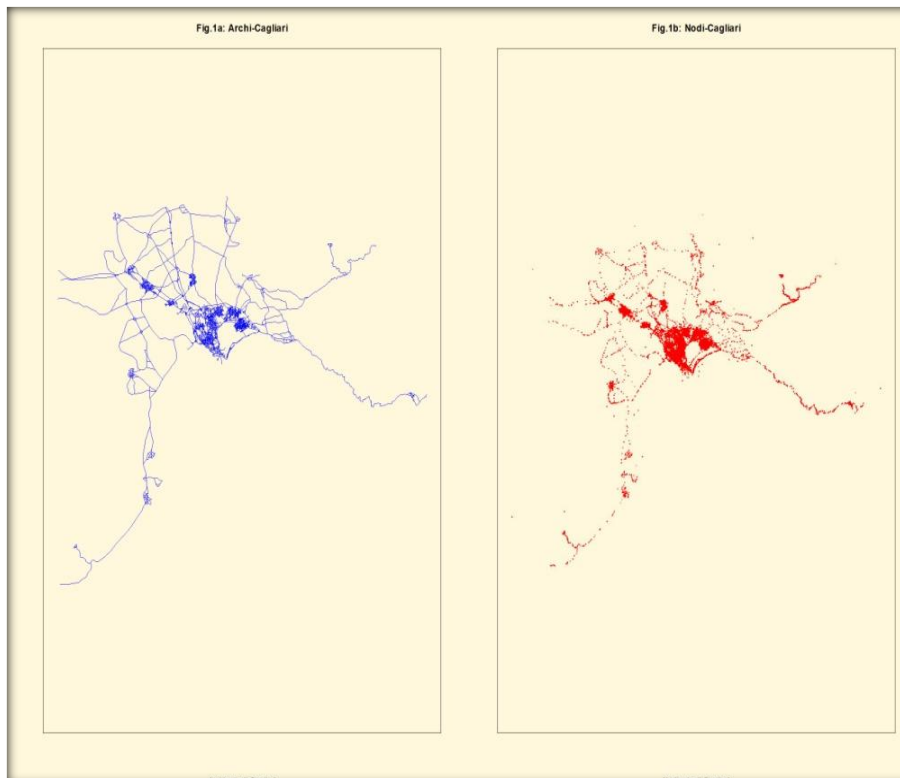
N	X	Y
610	1515334,00000	4342400,00000
611	1516827,05454	4344216,06397
612	1516868,00000	4344255,00000
613	1515029,00000	4343669,00000
665	1514958,00000	4343682,00000
729	1508418,00000	4342738,00000
730	1510793,00000	4342290,00000
731	1510793,00000	4342281,00000
732	1511704,00000	4342559,00000

Tabella 6: Alcune informazioni di alcuni archi della rete stradale

A	B	DISTANCE	SPEED	CAPACITY	LINKTYPE	Link
610	2273	0,12	30	400	5	610_2273
610	4744	0,09	3	400	5	610_4744
611	612	0,05	50	1600	3	611_612
611	4802	0,23	50	1600	3	611_4802
612	611	0,05	50	1600	3	612_611
612	12202	0,03	70	1300	3	612_12202

Nel file contenente le informazioni sugli archi della rete, sono riportate numerose informazioni relative a dati topologici dell'arco; nella Tabella 6 sono riportate solo alcune delle informazioni presenti quali il numero del nodo iniziale e finale dell'arco, la loro distanza, la capacità, il tipo di strada e un ultimo campo non numerico con l'indicazione di entrambi i nodi. Nel seguito vengono riportati i grafici dei nodi e della rete stradale di Cagliari riprodotti con il linguaggio di programmazione R. algoritmi di MM utilizzano dati provenienti da GPS e li integrano con informazioni relative alla risoluzione spaziale del territorio.

Figura 11: Rappresentazione degli archi e nodi della rete stradale della città di Cagliari



Con dati GPS non affetti da errore e mappe spaziali territoriali con un alta risoluzione non ci sarebbe la necessità di operare nessuna procedura preliminare all'analisi. Ma gli errori di rilevamento satellitari e una mappa territoriale non sempre precisa ed esaustiva rendono necessario il MM.

Gli algoritmi possono produrre soluzioni più o meno imprecise a causa dei problemi rilevati in precedenza, ovvero ad una scarsa qualità locale della mappa digitale, agli errori non eliminabili delle trasmissioni satellitari, alla tipologia territoriale o stradale (effetto tunnel) oppure all'imprecisione del ricevitore GPS. Le informazioni ottenute dal GPS possono essere condizionate da molti fattori, tra cui le condizioni del tempo, la posizione non corretta dello strumento per la decodifica delle trasmissioni satellitari, gli ostacoli ambientali, come grandi masse di acqua, alti rilievi o canyon naturali o artificiali, strade tra costruzioni alte o in generale comprese tra ostacoli, forti differenze qualitative fra marche e modelli diversi di ricevitori GPS, l'elevata congestione dell'etere che può disturbare il segnale.

La complessità del sistema stradale, il numero di incroci o rotonde la presenza di cavalcavia o tunnel, la presenza di foreste, di montagne o la densità di alte costruzioni può notevolmente degradare l'informazione relativa alla traiettoria seguita da un veicolo fornito di ricevitore satellitare. Questi fattori non possono essere eliminati ma i loro effetti possono essere ridotti o annullati dal *postprocessing*.

2.2 Tipologie dell'approccio geometrico

Le procedure di MM variano a seconda del tipo di problema che si vuole affrontare. Nel corso di quest'ultimo ventennio le tecniche di MM si sono evolute in funzione del cambiamento della richiesta dei servizi e

della crescente richiesta delle applicazioni dell'ITS (*Intelligent Transport System*), per l'utilizzo di metodologie diverse, per la nascita di una nuova casistica. In particolare la possibilità di rilevare grandissime quantità di dati e la facilità della loro gestione, grazie all'uso sempre più spinto della tecnologia informatica, ha permesso di migliorare alcune tecniche e di sperimentarne altre; l'utilizzo dei sistemi di rilevamento sempre più massivo e poco costoso e la possibilità di aggiornamento *friendly oriented*, ha aumentato la casistica di studio e aperto nuove frontiere. La richiesta dell'ITS su bus, o su veicoli di emergenza, per le informazioni stradali o per il *fleet management* ha migliorato il settore di applicazione del rilevamento *on line*; lo sviluppo tecnologico dei ricevitori satellitari e dei servizi satellitari ha permesso di porre l'attenzione sul miglioramento delle tecniche computazionali piuttosto che sull'eliminazione degli errori di trasmissione-ricezione. Lo sviluppo di settori di ricerca, come il *data mining* o le reti neurali hanno aperto nuovi scenari di sperimentazione e applicazione .

I metodi di *map matching* vengono generalmente classificati in approccio geometrico, topologico, probabilistico e di tipo avanzato. L'approccio di tipo geometrico utilizza informazioni geometriche riguardanti la mappa digitale, considerando solo il percorso seguito, cioè non prendono in considerazione la connessione esistente tra un arco ed un altro. Il più semplice approccio geometrico e quello di più facile implementazione è quello *point to point* (Bernstein and Kornhauser, 1996) che considera i punti GPS e i nodi della rete stradale più prossimi alle rilevazioni. La figura seguente mostra un esempio di tale approccio: la rete stradale è composta dai tratti B^1C , B^1B^0 , A^2A^1 , A^1A^0 . I punti rilevati sono P^0 fino a P^6 : i punti verranno agganciati ai nodi della rete più vicina.

Ma se il percorso effettivamente seguito è $B^0 - B^1$ quello che risulta dopo l'attribuzione è $B^0 - A^1 - A - B^1$.

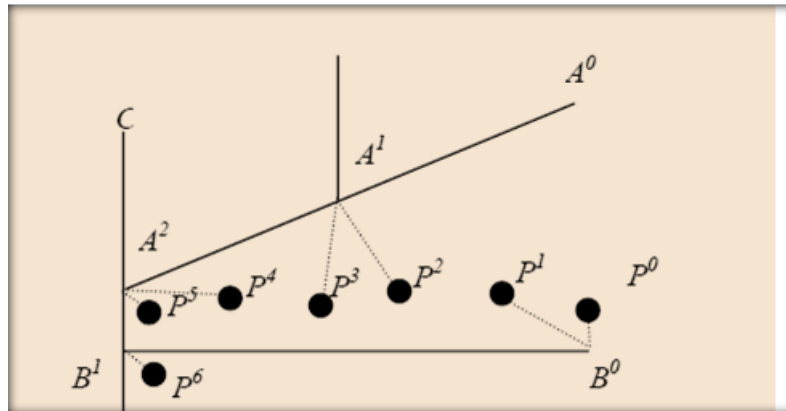


Figura 12: Un esempio di approccio geometrico *point to point*

Un altro approccio è quello conosciuto come *point to curve matching*, (White *et al.*, 2000) per cui i punti sono uniti all'arco di rete più vicino; ogni curva è composta da segmenti che sono a tratti lineari e la distanza è calcolata tra il punto e ognuno dei segmenti: il segmento posto ad una distanza minima rispetto agli altri è il prescelto per simulare il percorso del veicolo.

La Figura 13 mostra lo stesso esempio precedente ma con un risultato diverso: i punti P^4 P^6 si agganciano a tratti di strada non corretti e l'errore aumenta se la rete urbana è ad alta densità di strade e crocevia.

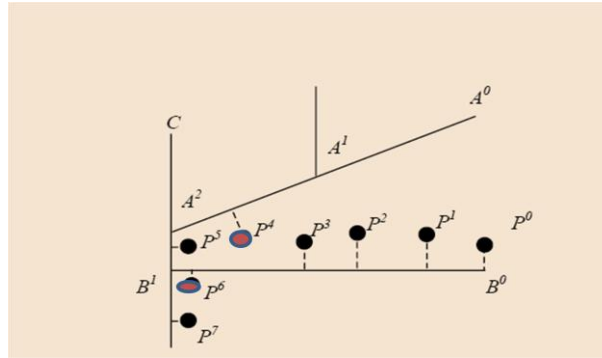


Figura 13: Un esempio di approccio geometrico *point to curve*

Per calcolare la distanza minima si fa uso della distanza perpendicolare tra il punto e l'arco di strada risolvendo le seguenti equazioni:

$$X = \frac{m(y_3 - y_1 + mx_1) + x_3}{m^2 + 1}$$

$$Y = y_1 + m(X - x_1)$$

$$D = \sqrt{(X - x_3)^2 + (Y - y_3)^2}$$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

con (x_1, y_1) , (x_2, y_2) , (x_3, y_3) le coordinate dei punti GPS P_1 , P_2 , P_3 e (X, Y) le coordinate del nodo A di rete come mostrato in Figura 14.

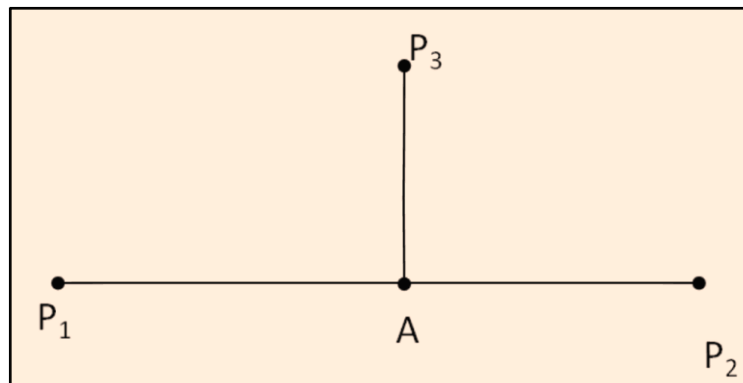


Figura 14: La distanza minima ad un arco

Un altro metodo geometrico è quello *curve to curve*, (White *et al.*, 2000; Phuyal, 2002) in cui si utilizza in prima battuta l'approccio puntuale per identificare i nodi candidati sulla rete; successivamente per ogni nodo candidato si costruiscono curve a tratti lineari dai percorsi che sono originati da quel nodo. Tale approccio condente poi di costruire il percorso GPS e confrontarlo utilizzando le distanze geometriche con i tratti di strada più vicini. La strada più prossima al percorso è presa come quella che effettivamente il veicolo ha attraversato. Anche in tale metodologia il confronto tra il percorso e la rete può portare in errore: come mostra la figura successiva se si confronta il percorso $P_1.P_2$ con i tratti di rete $A_1.A_2$ e $B_1.B_2$ e si confrontano le distanze si nota come l'attribuzione del percorso a di un tratto piuttosto che ad un altro tratto di strada non sempre è corretta. Se poi ci troviamo tra due strade parallele tra loro o in un incrocio, come nella Figura 15, l'ambiguità risulta essere maggiore.

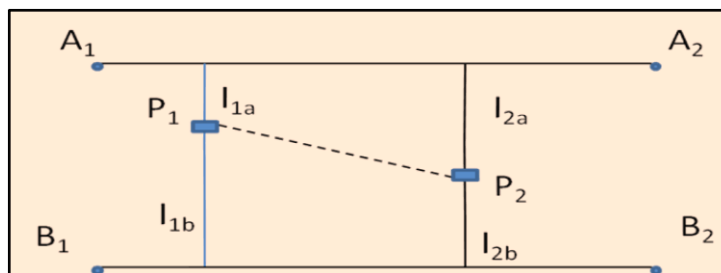


Figura 15: Distanza minima tra un punto e due strade parallele tra loro

La principale limitazione dell'approccio geometrico è che si ignora la sequenza dei punti nel tempo e la possibile connessione tra gli archi di rete; inoltre solo la ricerca del nodo più vicino rende la procedura veloce. Nielsen *et al.* (2004) sottolineano che il tasso di errore rilevato in un'indagine condotta su 3000 viaggi effettuati nell'area di Copenhagen associati a 300000 archi di rete è stato talmente elevato (33%) da non poter rendere reale nessuna applicazione.

2.3 L'approccio topologico

L'approccio topologico al MM usa le relazioni di connettività e di adiacenza tra le diverse entità considerate, ovvero punti, linee e poligoni, utilizzando la mappa topologica. Greenfeld (2002) propone un algoritmo topologico che fa uso solo delle informazioni sulle coordinate del percorso e non vengono prese in considerazione altre informazioni determinate dal ricevitore GPS, come la velocità e la direzione di avanzamento (*heading*). Il MM proposto si compone di due algoritmi: il primo per ricercare il punto iniziale del percorso, selezionando quel nodo della rete più vicino al primo punto GPS e definire successivamente tutti gli archi connessi a questo nodo. Il secondo algoritmo opera un'analisi topologica seguendo uno schema pesato ed è composto dai seguenti passi:

- ricava il successivo punto GPS
- definisce un segmento tra il punto precedente e il punto ottenuto nel precedente passo
- valuta la distanza e la direzione (Azimuth) del tratto GPS rispetto al segmento di strada A_i
- se il punto definito nel primo passo non si associa al segmento di strada A_i , se ne cerca un altro A_{i+1} (utilizzando lo stesso schema di prossimità e orientazione) connesso ad A_i sulla stessa strada o su di una adiacente

Questo algoritmo è però molto sensibile agli *outliers* ovvero a quei punti GPS che presentano posizioni estreme rispetto alla totalità delle osservazioni. Può capitare che l'angolo di intersezione tra il percorso e la

rete porti ad una errata identificazione del tratto di strada come si vedrà successivamente in dettaglio.

Molti algoritmi proposti utilizzano l'approccio geometrico per la ricerca del nodo iniziale e successivamente analizzano l'ultimo link con tutti percorsi successivi, anche se alcuni (Chung and Shalaby, 2005) prendono in considerazione anche quelli precedenti. Per la scelta del miglior link alcuni autori utilizzano la distanza minima euclidea tra i punti e le proiezioni ortogonali degli stessi sul link (distanza perpendicolare) o la distanza rispetto al punto iniziale o finale del tratto considerato. L'approccio topologico restituisce risultati migliori in termini di velocità computazionale e in accuratezza del percorso rispetto al metodo geometrico, in quanto prende in considerazione la sequenza completa dei punti e un limitato numero di link, ed è meno sensibile agli errori di misurazione e agli *outliers*.

Per aree urbane molto dense e quindi con un numero notevole di strade, operare un MM è molto più complesso in quanto la distanza media tra le strade è molto bassa. Per diminuire la probabilità di errore bisognerebbe utilizzare tutte le informazioni che provengono dal GPS e dalla mappa digitale: ad esempio utilizzare la velocità e l'*heading* del veicolo migliorerebbe l'identificazione della strada corretta. E' da sottolineare che, però, soprattutto negli incroci dove la velocità del veicolo è più bassa la misura dell' *heading* rilevata dal GPS tende ad essere instabile; l'utilizzo di un giroscopio migliora sensibilmente la qualità dell'informazione. il *Dead Reckoning* (DR) è un dispositivo inerziale che utilizza giroscopi, bussola e accelerometri e permette di migliorare l'informazione e di continuare la stima della localizzazione di una posizione anche quando non è possibile ricevere il segnale satellitare tramite il GPS. Tuttavia, utilizzare l'informazione sulla velocità per

definire la strada corretta può comportare svantaggi superiori agli eventuali vantaggi. Infatti l'addensamento di punti di rilevazione del veicolo molto concentrati in prossimità di una strada potrebbe essere causato ad esempio dalla presenza di un semaforo che obbliga il veicolo a fermarsi: in questa circostanza l'*heading* è altamente instabile può variare l'orientamento in maniera notevole.

Per selezionare i tratti di strada candidati alcuni autori hanno utilizzato una circonferenza circolare con un raggio fisso intorno alla posizione fissata per tutti i punti presi come riferimento nei tempi successivi. Questa tecnica però, laddove le aree sono caratterizzate da una densità elevata di strade ed incroci, può provocare errori di posizionamento: per questo sono preferibili regioni che cambiano dinamicamente. Pyo et al. (2001) selezionano una regione ellittica che varia in base agli errori del GPS. Un altro inconveniente che può verificarsi è che la regione di confidenza non contenga la linea centrale della strada, soprattutto quando la larghezza della strada è elevata.

2.4 Le procedure avanzate per il *map matching*

Nel tempo i metodi deterministici si sono evoluti in algoritmi probabilistici: il maggior vantaggio di tali metodi è che non si assume che il veicolo sia sulla rete; si definisce una regione di confidenza GPS ellittica o rettangolare intorno al punto all'interno della quale deve essere trovata la posizione del veicolo. Se nella regione si trova un segmento di strada allora si opera la corrispondenza e le coordinate della strada sono utilizzate per calcolare la posizione successiva del veicolo. Se una o più strade sono candidate per una corrispondenza tra il percorso e la reale

traiettoria si opera un esame delle connettività per determinare la posizione più probabile; alla migliore corrispondenza si associa una probabilità.

Zhao (1997) introduce questa tecnica stabilendo che se nella regione di confidenza (o di errore) ricadono diversi segmenti di rete, allora il segmento candidato viene scelto in base a criteri di connettività e di vicinanza. La grandezza della regione di confidenza deriva dall'errore di varianza. Oichieng *et al.* (2004) studiano tutti i percorsi che ricadono nella regione in base alla distanza, alla direzione e alla connettività dei precedenti link, e talvolta anche rispetto alla velocità.

Gli algoritmi di MM avanzati utilizzano concetti più ricercati come il Filtro Kalman, la teoria matematica dell'evidenza di Dempster-Shafer, il *particle filter*, il modello interattivo multiplo, il modello *fuzzy logic* o le applicazioni di inferenza bayesiana. Nelle appendici sono riportate alcune delle teorie citate. Kim *et al.* (2000) propongono l'utilizzo del filtro Kalman esteso per minimizzare l'errore di proiezione ortogonale che si commette quando il punto GPS è lungo la traiettoria (*along track error*). Anche in questo caso la prestazione del filtro dipende dai dati della rete geografica e, l'applicazione del metodo geometrico per l'identificazione del link può portare ad un link non corretto e ad una successiva propagazione dell'errore.

Gustafsson *et al.* (2002) sviluppano una metodologia legata ai problemi di posizione, navigazione e di allineamento utilizzando come filtro una stima ricorsiva bayesiana (RBE). Una delle applicazioni di questo metodo è quella di ricercare la posizione del veicolo utilizzando la velocità di spostamento. Gli autori sottolineano come un'errata definizione della posizione iniziale dell'ordine di un chilometro viene migliorata di un

metro applicando la stima ricorsiva. Il punto iniziale viene stabilito direttamente dal guidatore oppure calcolato mediante il ricevitore GPS.

Cui and Ge (2003) propongono una soluzione laddove il segnale GPS è ostacolato dal cosiddetto effetto canyon, approssimando il percorso con parti di curve come linee, archi e polinomiali. Per identificare il corretto segmento di rete in corrispondenza dell'incrocio utilizzano un algoritmo probabilistico che viene ulteriormente integrato con un filtro Kalman esteso al fine di stimare la posizione del veicolo nell'incrocio ed identificare la corretta strada tra tutte quelle che si incontrano nell'incrocio stesso. Questo algoritmo prevede la visibilità di soltanto due satelliti e quindi la ricerca della posizione in un'area altamente densa migliora. Se non sono disponibili almeno due satelliti l'algoritmo non raggiunge i risultati sperati.

Fu *et al.* (2004) propongono un algoritmo ibrido per analizzare la geometria della rete; per identificare il miglior link tra diversi candidati utilizzano un modello *fuzzy logic* che però risulta essere sensibile ai rumori di misurazione. Gli autori prendono in considerazione la distanza minima tra il punto e il segmento e la differenza tra la direzione del veicolo e la direzione del segmento di strada. Il risultato del modello applicato è la possibilità di associare il punto al segmento. Tuttavia in tale approccio non si considera la circostanza per cui a basse velocità la misura dell'*heading* assume valori non attendibili; inoltre, non prendendo in considerazione la traiettoria precedente al punto preso in considerazione, esiste un'alta probabilità di selezionare un segmento di strada errato.

Le limitazioni di tale modello vengono superate da Quddus *et al.* (2007) che utilizzano una serie di variabili nelle regole fuzzy quali la

velocità del veicolo, la connettività dei link lungo tutto il percorso, la qualità della soluzione e la posizione di un punto rispetto al candidato link.

Nielsen *et al.* (2004) propongono un algoritmo che non utilizza la regione di errore intorno al punto ma sfrutta quello dei cammini minimi di Dijkstra: iniziando dal nodo origine, la strada è sviluppata aggiungendo i nodi finali di tutti i link uscenti del nodo corrente del set dei nodi presi in considerazione. Il nodo successivo preso in considerazione è quello che sarà ottenuto nel minor tempo possibile, iniziando dall'ultimo nodo preso in considerazione. Questo procedimento oltre a comportare un elevato errore laddove l'assegnazione iniziale è errata, non garantisce la soluzione ottimale così come l'algoritmo di Dijkstra.

Pyo *et al.* (2001) introducono la tecnica dell'ipotesi multipla (MHT) per definire la regione di confidenza e l'identificazione non di una sola strada, ma di una serie di percorsi facenti parte di un set di strade-candidate, e per ognuno di questi percorsi si associa un punteggio basato sulle distanze perpendicolari: il miglior candidato è quello individuato alla fine del percorso. Il numero dei candidati non può essere illimitato né comunque elevato per una praticità computazionale: Pyo non definisce nessun criterio di sfoltimento. Il set delle ipotesi è definito sull'analisi topologica della mappa digitale, e sulle cosiddette pseudo misurazioni, definite come la proiezione dei punti rilevati dal GPS sulla rete, ovvero posizione e *heading*. Queste vengono generate per tutti i link all'interno della regione di errore; insieme all'analisi topologica esse sono utilizzate per definire sia il set delle ipotesi che la loro probabilità per ogni informazione del GPS/DR. Se una probabilità è più bassa di un valore di soglia definito, allora l'ipotesi viene eliminata; anche se il rapporto tra la probabilità di una ipotesi e la più alta probabilità risulta essere inferiore rispetto ad un altro valore di soglia l'ipotesi viene scartata. Lo svantaggio

di questa metodologia è che non è definito alcun metodo per l'identificazione del punto iniziale e di conseguenza i risultati risentono di un eventuale errata identificazione della prima associazione.

Marchal et al. (2005) definiscono un algoritmo di ricerca topologica utilizzando la tecnica dell'ipotesi multipla: il set iniziale è determinato dagli m nodi più vicini al punto iniziale e creando un singolo percorso per ognuno dei percorsi incidenti. Successivamente si valuta se il punto GPS considerato può essere associato all'ultimo link della strada o se è presente un incrocio. In quest'ultimo caso viene creata una nuova strada candidata per ogni link successivo al link corrente. Il nuovo link è aggiunto alla fine della strada. Ad ogni nuovo candidato viene assegnato un punteggio e inserito in un insieme di strade candidate; il numero dei candidati è limitato e viene definito eliminando tutti quelli che raggiungono un punteggio inferiore rispetto a quelli che sono stati utilizzati con l'ultimo punto GPS. Laddove i dati GPS siano su larga scala e la topologia della rete sia ad alta risoluzione, l'algoritmo dimostra essere più preciso e veloce di quello di Pyo et al. che non definiscono un limite al numero dei percorsi candidati.

Ochieng *et al.* (2004) sviluppano un algoritmo probabilistico per cui la regione di errore ellittica è definita solo se il veicolo attraversa un incrocio; questo rende l'algoritmo più veloce. Inoltre sviluppano una serie di criteri basati su studi empirici per identificare manovre del veicolo in un incrocio: questo può essere di aiuto per riconoscere un cambiamento di strada del veicolo. L'algoritmo prende in considerazione anche la inattendibilità dell'*heading* quando il veicolo viaggia a basse velocità.

Gli algoritmi di MM si possono differenziare anche rispetto all'intervallo di rilevazione del GPS per il quale si associano tassi di

campionamento più o meno alti: Yuan *et al.* (2009) propongono un algoritmo come in Pyo (2001) e in Marchal (2005) ma adatto per tassi di campionamento GPS bassi. In ultimo ricordiamo una ulteriore possibile suddivisione tra algoritmi locali e globali: i primi seguono una strategia di risoluzione del problema metaeuristica ovvero ricercano il punto o l'intorno locale ottimo, sulla base della similarità della distanza e dell'orientazione; i secondi cercano di trovare una traiettoria la più vicino possibile a quella osservata tra tutte quelle disponibili nella rete presa in considerazione.

La maggiore limitazione caratteristica degli algoritmi di MM è relativa all'identificazione del punto iniziale e quindi all'associazione di un tratto del percorso ad una strada della rete, soprattutto se è complessa, in cui sono presenti incroci a quadrifoglio, cavalcavia o svincoli. Le tecniche maggiormente usate sono quelle di definire delle regioni di errore circolari o ellissoidali in cui ricada il punto di rivelazione GPS; se però il punto non rientra in questa regione circoscritta, l'identificazione non avviene o risulta essere errata e questo errore si ripercuote in tutte le successive fasi della procedura.

Per ridurre gli errori dovuti alla complessità di rete o ali errori di misurazione GPS, la comunità scientifica che si occupa dell'argomento ha ritenuto di integrare le tecniche di MM con informazioni topologiche, dalle più semplici informazioni di connettività a parametri più complessi. Ciononostante, i risultati sono stati sempre a di sotto delle aspettative; Quddus *et al.* (2007) attribuiscono questa bassa performance alle informazioni spesso carenti delle mappe digitali.

2.5 La valutazione dei risultati

Per valutare il risultato dell'algoritmo applicato al problema, ovvero se il risultato del MM è corretto, si paragonano i risultati ottenuti dalla procedura con la traiettoria reale. Se il percorso GPS determinato dal riferimento della traiettoria del veicolo coincide con quella ottenuta con le procedure di MM, allora si sarà identificato il percorso correttamente.

E' necessario operare una valutazione delle performance degli algoritmi applicati; si nota però che in molti studi ed applicazioni questo passaggio non sempre è presente ed in particolare spesso la metodologia non viene supportata dall'analisi della accuratezza orizzontale. Quddus *et al.* (2006) hanno proposto di stimare la performance ipotizzando che al tempo t il punto rilevato sia $G(x,y)$, $M(x_2,y_2)$ il punto stimato dal MM, mentre la posizione reale del veicolo sia $A(x_1,y_1)$. Si avrà, allora al tempo t una precisione orizzontale MA pari a :

$$MA = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Dove la differenza (x_2-x_1) rappresenta l'accuratezza della longitudine e la differenza (y_2-y_1) quella della latitudine. Nella Figura 16 si riporta la rappresentazione schematica.

Questa misura può essere applicata in tutti i tempi di cui è composto il percorso. Le misure descrittive (media, deviazione standard) associate possono essere usate per determinare il risultato qualitativo del MM. Definito l'angolo θ tra la traiettoria I (*along track*) e la rotta I (il *cross-track*) è possibile calcolare la loro precisione come segue:

$$MC = MA \cos \theta$$

$$AC = MA \sin \theta$$

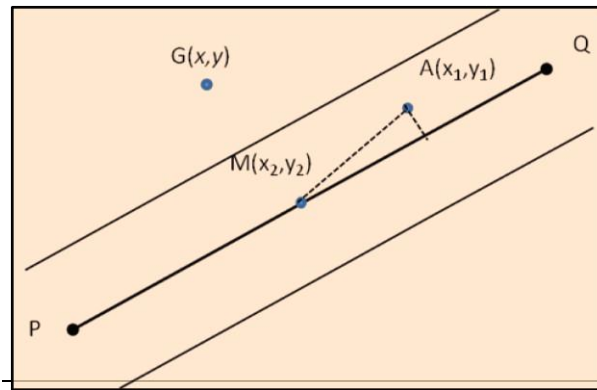


Figura 16: Posizione rilevata, stimata e reale di un veicolo

Nella Tabella 7 sono riepilogati alcuni algoritmi di MM: nella prima colonna sono indicati i nomi degli autori e l'anno di pubblicazione dell'articolo correlato; la seconda colonna è indicato l'approccio utilizzato ed infine alcune caratteristiche dell'algoritmo.

Bernstein and Kornhauser (1996) Kim et al. (1996) White et al. (2000)	Point to point	Non usa informazioni "storiche"; sensibile al modo in cui si propone il network	Dichieng et al. (2004)	Approccio probabilistico	Utilizzano la regione dell'errore. Negli incroci poco efficiente. Veloce e robusto a basse velocità.
Bernstein and Kornhauser (1996) White et al. (2000)	Point to curve	Mancanza di informazioni topologiche determina assegnazione incorretta	Fu et al. (2004)	Fuzzy logic	Non considera l'errore dell'angolo a basse velocità e le informazioni storiche.
Taylor et al. (2001) White et al. (2000)	Point to curve con informazioni di angolazione	Inaccurato a basse velocità	Nielsen et al. (2004)	Dijkstra	Non garantisce la ricerca del percorso ottimale.
Bernstein and Kornhauser (1996) Taylor et al. (2001) White et al. (2000)	Point to curve	Sensibile agli outliers e con risultati indesiderabili	Chung and Shalaby (2005)		L'ultimo link con tutte le strade precedenti
Bernstein and Kornhauser (1996) White et al. (2000)	Point to curve e topologico	Un cattivo punto iniziale si ripercuote su tutto l'algoritmo	Pyo et al. (2001) Marchal (2005)	Approccio probabilistico	Tecniche di ipotesi multipla e utilizzo della regione di errore; filtro Kalman
Greenfield (2002)	Approccio topologico	Cattiva performance in presenza di intersezioni di strade	Quddus et al. (2006)	Fuzzy logic	Utilizza la regione di errore e variabili aggiuntive. la velocità, la connettività sul link, la qualità dell'osservazione, la posizione del link candidate
Quddus et al. (2003)	Filtro Kalman esteso	Si utilizza un GPS/DR e si considerano informazioni sulla posizione, velocità e tempo. Efficiente nelle intersezioni e giunzioni delle strade	Yuang (2009)	Geometrico, topologico, voting	Con un'analisi spaziale e temporale si attribuiscono votazioni ai percorsi definendo il miglior percorso. Per basso tasso di campionamento.
Cui and Ge (2003)	Approccio probabilistico, Filtro Kalman	Ha bisogno di due satelliti disponibili. Diminuisce problemi Canyon			

Tabella 7: Alcuni algoritmi di map matching e loro caratteristiche

3. L'implementazione di alcune tipologie di *map matching*

In questo capitolo si descrivono alcune procedure di *map matching* che utilizzano sia l'approccio topologico che altri metodi cosiddetti avanzati. Il primo algoritmo proposto si basa sull'analisi topologica ed utilizza come informazioni del veicolo solo le coordinate GPS e nessuna ulteriore informazione della posizione del veicolo; della mappa digitale prende in considerazione le informazioni di connettività, adiacenza e prossimità. Ad ogni strada candidata per l'associazione al percorso GPS, viene assegnato un punteggio: la migliore sarà quella a cui è associato il punteggio maggiore. Questo approccio viene migliorato dalla proposta di Qudus *et al.*(2006b) che utilizza anche le informazioni relative alla velocità del veicolo, la posizione del veicolo rispetto alle strade candidate e la direzione del percorso. Sono state sviluppate anche procedure per la ricerca del percorso più breve, laddove si richieda di identificare il percorso o di compiere una scelta alternativa in base alle informazioni in tempo reale: si descrivono, quindi, alcune proposte di ricerca di soluzioni ottimali mediante strategie di ricerca euristica. L'approccio probabilistico prevede la creazione di intervalli di confidenza grazie ai quali vengono valutate quelle strade candidate che ricadono in tali intervalli: vengono descritte due proposte di procedure di *map matching* e come si possono definire le regioni dell'errore. Infine, prendendo spunto da una proposta operativa, si descrive come è possibile utilizzare delle informazioni di tipo qualitativo utilizzando la logica e i modelli *fuzzy*: il processo di identificazione della corretta strada percorsa da un veicolo viene visto

come un processo di decisione in cui il grado di ambiguità viene espresso da un modello matematico, che utilizza regole deduttive sia in entrata che in uscita.

3.1 Un esempio di algoritmo topologico

Di seguito viene descritto l' algoritmo di MM con approccio topologico proposto da J. S. Greenfeld (2002), da cui nel tempo si sono ispirate altre tecniche metodologiche, come quella proposta da Quddus et al. (2003).

La procedura di MM è composta da due algoritmi separati: il primo è detto *Initial Mapping* ed è utilizzato per definire il punto iniziale, il secondo detto *Map* ed è la parte principale della procedura successiva all'applicazione del primo algoritmo. Esso usa informazioni topologiche per unire i punti P^t agli archi A_i della mappa digitale N .

L'algoritmo iniziale viene applicato in diverse occasioni: quando si riceve il primo punto P^0 dal GPS, o quando la distanza temporale tra i punti P^t e P^{t+1} è superiore ad un livello di tolleranza definito a priori. Questo è il caso in cui ci sia una perdita del segnale e quindi la necessità di ricalcolare il punto iniziale. Si usa infine quando l'algoritmo principale non è riuscito a raggiungere alcun risultato e quindi la procedura riparte nuovamente.

Il primo algoritmo, per la ricerca del punto iniziale, individua il nodo A^0 della strada più vicina al punto P^0 sfruttando informazioni unicamente geometriche; definisce tutti gli archi in N che sono connessi ad A^0 . Quando è disponibile il punto successivo inserisce il punto P^0 in uno degli archi selezionati.

L'algoritmo principale è così strutturato: acquisisce il punto P^t ; definisce un segmento tra due punti consecutivi P^{t-1} e P^t ; valuta la distanza e la

direzione (o azimuth) della linea GPS creata rispetto al tratto di strada A^i ; se il punto successivo P^t non può essere inserito sul segmento A^i allora si troverà un altro segmento A^{i+1} connesso ad A^i o nei suoi pressi. Anche l'arco A^{i+1} è selezionato in base ai criteri di prossimità ed orientamento.

3.1.1 La ricerca della similarità e dell'orientamento

Di seguito viene descritto l' algoritmo di MM topologico proposta da J. S. Greenfeld (2002), da cui nel tempo si sono ispirate altre tecniche metodologiche, come quella proposta da Quddus *et al.* (2003).

La procedura di MM è composta da due algoritmi separati: il primo è detto *Initial Mapping* ed è utilizzato per definire il punto iniziale, il secondo detto *Map* ed è la parte principale della procedura successiva all'applicazione del primo algoritmo. Esso usa informazioni topologiche per unire i punti P^t agli archi A_i della mappa digitale N .

L'algoritmo iniziale viene applicato in diverse occasioni: quando si riceve il primo punto P^0 dal GPS, o quando la distanza temporale tra i punti P^t e P^{t+1} è superiore ad un livello di tolleranza definito a priori. Questo è il caso in cui ci sia una perdita del segnale e quindi la necessità di ricalcolare il punto iniziale. Si usa infine quando l'algoritmo principale non è riuscito a raggiungere alcun risultato e quindi la procedura riparte nuovamente.

Il primo algoritmo, per la ricerca del punto iniziale, individua il nodo A^0 della strada più vicina al punto P^0 sfruttando informazioni unicamente geometriche; definisce tutti gli archi in N che sono connessi ad A^0 . Quando è disponibile il punto successivo inserisce il punto P^0 in uno degli archi selezionati.

L'algoritmo principale è così strutturato: acquisisce il punto P^t ; definisce un segmento tra due punti consecutivi P^{t-1} e P^t ; valuta la distanza e la direzione (o azimuth) della linea GPS creata rispetto al tratto di strada A^i ; se il punto successivo P^t non può essere inserito sul segmento A^i allora si troverà un altro segmento A^{i+1} connesso ad A^i o nei suoi pressi. Anche l'arco A^{i+1} è selezionato in base ai criteri di prossimità ed orientamento.

Possono essere utilizzati due metodi per determinare la prossimità di un punto P^t all'arco A^i . Il primo è calcolare la più piccola distanza perpendicolare tra il punto P^t e l'arco A^i . Questo metodo però non dà la certezza che il punto sia correttamente posizionato sull'arco A^i , ma che tale arco è un candidato. Il secondo metodo è determinare se il segmento tra P^t e P^{t-1} e l'arco stesso si intersecano e se hanno la stessa direzione: in tale circostanza il posizionamento sull'arco è corretto. Ma se l'intersezione avviene con un angolo prossimo ai 90° , ovvero se i due segmenti tendono ad essere perpendicolari, allora il posizionamento sull'arco sarà errato, in quanto i due segmenti divergono. Se l'analisi viene condotta sui loro prolungamenti il criterio di intersezione non è valido.

Per definire l'orientamento o Azimuth si confrontano le direzioni del segmento passante per i due punti GPS e quello passante per due nodi di strade consecutive. Si preferisce utilizzare la misura in Azimuth in quanto questa misura è compresa tra 0 e 2π e definisce senza ambiguità linee parallele o perpendicolari. Se questa misura non può essere ottenuta dalle informazioni rilasciate dal GPS dovrà essere definita come segue:

$$Az_{P^1P^2} = \tan^{-1} \frac{x_2 - x_1}{y_2 - y_1}$$

se $y_2 - y_1$ non è una quantità positiva allora si aggiungerà π ; se Az è negativo si aggiungerà 2π .

Un sistema di pesi valuta quali archi sono i candidati per un punto P^i mediante il calcolo di un punteggio di verosimiglianza basato su tre indicatori diversi. Il primo con cui si calcola il punteggio W da associare agli archi candidati si basa sul grado di parallelismo tra il segmento GPS passante tra due punti contigui e l'arco di strada, definito come W_{AZ} ; il secondo definisce la prossimità del punto all'arco di strada, W_D ; il terzo criterio definito come W_I , stabilisce l'eventuale angolo di intersezione tra i due segmenti. Per calcolare i vari pesi l'autore propone le seguenti formulazioni:

$$W_{AZ} = C_{AZ} \cos^{n_{AZ}}(\Delta AZ)$$

$$W_D = C_D - aD^{n_D}$$

$$W_I = C_I \cos^{n_I}(\Delta AZ)$$

con ΔAZ costante e massimo punteggio che rappresenta il grado di parallelismo tra il segmento del percorso GPS e l'arco di strada della mappa digitale. Se allora C_{AZ} è maggiore di C_D si assume che la similarità nell'orientamento dei due segmenti è più importante rispetto alla prossimità dei due segmenti confrontati; attribuendo un valore maggiore a n_D rispetto a n_{AZ} si avrà che il peso totale W diminuirà più rapidamente quanto più il punto GPS è più lontano dall'arco A^i rispetto al decremento nel peso quanto più l'orientamento delle linee divergono l'una dall'altra. L'uso del coseno e l'assegnazione di un numero dispari ad n_i non solo decrementa il peso all'aumentare di ΔAZ ma permette di introdurre un peso negativo se ΔAZ è maggiore di 90° .

Nel seguito sono indicati alcuni esempi di come varia il peso complessivo, e quindi la regola di decisione. Nella Tabella 8 sono stati calcolati i diversi pesi e quello totale: sono stati utilizzati i parametri $C(i)$ e a usati da Greenfeld nella sua applicazione ad un caso reale. Si ipotizza

che le costanti $C(i)$ siano uguali per tutte le equazioni così come i parametri $n(i)$; l'angolo di intersezione tra il segmento GPS e il tratto di strada sia 10° , trasformati in radianti. All'aumentare di n_D sia il peso W_D che W diminuiranno drasticamente. Inoltre, una stessa variazione di n_I produce in W_I una diminuzione, anche se poco influente sul peso totale. (Tabella 9).

Tabella 8: Esempio di variazione dei pesi in funzione dei parametri

	$C(i)$	a	$n(i)$	$\cos(\delta)$	δ	Distanza	$W(i)$
W_{AZ}	10		5	0,98	0,17		9,26
W_D	10	0,1	5			2	6,80
W_I	10		5	0,98	0,17		9,26
W						Peso totale	25,33

Tabella 9: Esempio di variazione dei pesi in funzione dei parametri

	$C(i)$	a	$n(i)$	$\cos(\delta)$	δ	Distanza	$W(i)$
W_{AZ}	10		5	0,98	0,17		9,26
W_D	10	0,1	10			2	-92,4
W_I	10		10	0,98	0,17		8,58
W						Peso totale	-74,56

Se si aumenta la distanza D tra il punto GPS e l'arco il peso W_D subirà un forte decremento così come il peso complessivo W (Tabella 10).

Tabella 10: Esempio di variazione dei pesi in funzione dei parametri

	$C(i)$	a	$n(i)$	$\cos(\delta)$	δ	Distanza	$W(i)$
W_{AZ}	10		5	0,98	0,17		9,26
W_D	10	0,1	10			4	-104847,6
W_I	10		10	0,98	0,17		8,58
W						Peso totale	-104829,8

Se invece ΔAZ aumenta passando da 10° a 60° il peso di W_{AZ} e di W_I diminuisce così come W .

Tabella 11: Esempio di variazione dei pesi in funzione dei parametri

	$C(i)$	a	$n(i)$	$\cos(\delta)$	δ	Distanza	$W(i)$
W_{AZ}	10		5	0,5	1,05		0,31
W_D	10	0,1	5			2	6,8
W_I	10		10	0,5	1,05		0,01
W					Peso totale		7,12

Se l'angolo è di 90° , quindi il tratto del percorso GPS e il segmento della rete sono perpendicolari, i due pesi sono praticamente nulli mentre l'indicatore complessivo W raggiunge il minimo valore (Tabella 11).

Laddove un punto GPS è un *outlier*, o comunque un'osservazione che si colloca completamente fuori da un ipotetico percorso, allora la procedura prevede nuovamente l'utilizzazione del primo algoritmo per la ricerca del primo punto del percorso (*Inizial Mapping*). È possibile collocare un punto P_i su un arco A_i anche analizzando la direzione dei punti che formano il tratto del percorso precedente a P_i . Se però la direzione è molto diversa da quella del tratto precedente è possibile analizzare la direzione del tratto successivo a P_i . Poiché il cambiamento di direzione può dipendere dalla perdita del segnale, il periodo di attesa serve all'algoritmo per verificare se la direzione del tratto precedente sia compatibile con quella del tratto successivo al punto P_i . Analizzando i tratti precedenti e successivi al punto è possibile decidere se esso sia un *outlier* o meno. Inoltre, è possibile confrontare la distanza tra il punto GPS e il nodo finale dell'ultimo arco identificato: se la distanza è

maggiore di quella considerata come una ragionevole distanza di viaggio, allora P_i sarà associato all'arco A_i .

Come abbiamo visto in precedenza, la velocità del veicolo riveste un ruolo importante per la rilevazione dei punti GPS: se la velocità è elevata è possibile che non si riesca a collocare il percorso un arco di strada già associato al percorso in precedenza. La soluzione può ancora essere l'utilizzo del primo algoritmo per poter ridefinire il punto iniziale; anche se le informazioni precedenti saranno perse e bisognerà nuovamente ricollocare il percorso GPS con l'intero percorso reale. Se il punto P_i non può associarsi l'arco A_i o a nessun altro arco successivo si troverà la soluzione tra tutti gli archi che sono connessi al nodo finale di tutti gli archi possibili: non si considerano soltanto gli archi connessi al nodo finale dell'ultimo arco preso in considerazione, ma tutti gli altri archi successivi.

3.1.2 Una procedura in presenza di incroci

Quddus *et al.* (2003) dimostrano però cosa accade se si presenta un caso come nella figura precedente: si immagini che il percorso GPS sia composto dai punti P_i con i che varia da 1 a 4, e i

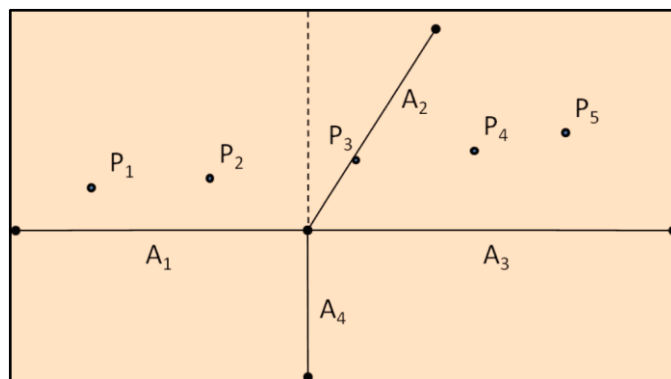


Figura 17: Un caso particolare di errata attribuzione

tratti di strada A_i con i che varia da 1 a 4. Si prenda in considerazione il punto GPS P_3 , le sue distanze dagli archi $D_1 = 40, D_2 = 0, D_3 = 40, D_4 = 2$ e gli angoli, espressi in gradi, con i diversi tratti di strada pari rispettivamente a $25^\circ, 60^\circ, 40^\circ, 155^\circ$ e si calcolino i diversi pesi totali W rispetto ai quattro archi di strada per il punto P_3 (Tabella 12, 13, 14) :

Tabella 12: Esempio di variazione dei pesi in funzione dei parametri

$A_1=A_3$	C	a	n	cos	Δ	D	
Waz	10		5	0,91	0,44		0
Wd	10	0,1	1,2			40	1,63
Wi	0		10	0,91	0,44		0
W							7,75

Tabella 13: Esempio di variazione dei pesi in funzione dei parametri

A_2	C	a	n	cos	Δ	D	
Waz	10		5	0,50	1,05		0
Wd	10	0,1	1,2			0,0	10,0
Wi	0		10	0,5	1,05		0
W							10,3

Tabella 14: Esempio di variazione dei pesi in funzione dei parametri

A_4	C	a	n	cos	Δ	D	
Waz	10		5	-0,91	2,7		0
Wd	10	0,1	1,2			2,0	9,78
Wi	0		10,0	-0,91	2,7		0
W							3,65

Il valore maggiore di W corrisponde a quello dell'arco A_2 che sarà il candidato per il punto P_3 . Ma questa associazione non è corretta. Quindi la metodologia non è robusta laddove si è in presenza di intersezioni tra strade.

Quddus *et al.* utilizzano la filosofia di fondo di Greenfeld, ma apportano alcune modifiche che migliorano notevolmente il *matching*. Dal punto di vista strumentale utilizzano un GPS-DR che permette di migliorare l'informazione e di stimare la localizzazione del veicolo anche in assenza di segnale. Inoltre, propongono che il punto iniziale sia quello rilevato dal GPS-DR più vicino alla strada e le informazioni relative alla velocità del veicolo. Inoltre il peso dell'equazione dell'orientamento e dell'*heading* W_H è definita in tal modo:

$$\begin{aligned}
 W_H &= A_H \cos(\Delta\beta') \quad \text{con } \Delta\beta = \beta - \beta' \\
 \Delta\beta' &= \Delta\beta \quad \text{se } -180^\circ \leq \Delta\beta \leq 180^\circ \\
 \Delta\beta' &= 360^\circ - \Delta\beta \quad \text{se } \Delta\beta > 180^\circ \\
 \Delta\beta' &= 360^\circ + \Delta\beta \quad \text{se } \Delta\beta < -180^\circ
 \end{aligned}$$

Con $A_H > 0$. Così, più il valore $\Delta\beta$ è basso più è alta la probabilità che il tratto di strada candidato sia corretto, ovvero maggiore risulta essere $\Delta\beta'$ minore sarà il valore di W_H . La funzione coseno permette di ottenere anche valori negativi, se $\Delta\beta'$ è maggiore di 90° .

Per definire il peso della prossimità del punto ad un tratto di strada si calcola la distanza perpendicolare tra il punto P_i e i nodi del tratto stesso. Dato un punto $P(x_3, y_3)$ e i nodi $A(x_1, y_1)$ e $B(x_2, y_2)$ la distanza perpendicolare è ottenuta come:

$$D = \frac{x_3(y_1 - y_2) - y_3(x_1 - x_2) + (x_1 y_2 - x_2 y_1)}{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}$$

Il peso della prossimità sarà quindi dato da:

$$W_{PD} = \frac{A_p}{D}$$

con $A_p > 0$ che è artefice della variazione e del controllo di W_{PD} . Maggiore è la distanza perpendicolare maggiore sarà il peso, ovvero P_i sarà più vicino all'arco e maggiore sarà la probabilità che il punto possa essere associato a quell'arco di strada. La distanza può essere calcolata anche prendendo in considerazione l'angolo che si forma tra l'arco di strada e il segmento passante per due punti GPS: più piccolo è l'angolo azimutale θ maggiore sarà la prossimità; se i segmenti presi in considerazione non si intersecano il punteggio varrà 0.

In tal caso se esiste intersezione si avrà la seguente formulazione del peso di prossimità:

$$W_{PI} = A_p \cos(\theta)$$

Per definire l'ultimo peso W_{RP} relativo alla posizione del punto rispetto al tratto di strada, si prendono in considerazione gli angoli che si formano tra il segmento passante per due punti consecutivi GPS e i tratti di strada da cui si vuole conoscere la distanza. Maggiore è l'angolo, più bassa sarà la probabilità che quel tratto di strada sia corretto per la corrispondenza:

$$W_{RP} = A_{RP} \cos(\alpha)$$

Con $\alpha \leq 180^\circ$ e $A_{RP} > 0$.

Per definire i parametri di W_H e W_{RP} si definiscono le seguenti equazioni:

$$A_H = aA_p$$

$$A_{RP} = bA_p$$

$$\text{con } b > a > 1$$

$$A_p > 0$$

dove a e b dipendono dalla performance dei sensori e dalla topologia della rete.

Infine, il punteggio del peso totale sarà calcolato rispetto al punto, per ogni link:

$$TW = W_H + (W_{PD} + W_{PI}) + W_{RP}$$

Si confrontino, ora, i risultati dell'esempio precedente applicando la metodologia di Quddus *et al.* Siano assegnati gli stessi parametri dell'esempio precedente e i seguenti valori per i parametri definiti:

$$C_{AZ} = 10; n_{AZ} = 5, C_D = 10, n_D = 1,2 C_I = 10, n_{AZ} = 10$$

$$D_1 = 40, D_2 = 1, D_3 = 40, D_4 = 2$$

$$\Delta\beta'_1 = 10^\circ; \Delta\beta'_2 = 75^\circ; \Delta\beta'_3 = 10^\circ; \Delta\beta'_4 = 100^\circ$$

$$\theta_1 = -; \theta_{12} = 75^\circ; \theta_3 = -; \theta_4 = -$$

$$\alpha_1 = 95^\circ; \alpha_2 = 0^\circ; \alpha_3 = 85^\circ; \alpha_4 = 175^\circ$$

$$a = 5, b = 2, A_p = 10$$

I TW calcolati per ogni tratto di strada $A1, A2, A3, A4$ per il punto P_3 saranno rispettivamente 47.74, 45.53, 51.23 e -23.60. Ciò comporta la scelta come miglior candidato per operare il *matching* con P_3 della strada A_3 . Il risultato è diverso da quello ottenuto con la metodologia di Greenfeld, è risulta inoltre essere corretto questo tipo di MM.

Il punto iniziale per la procedura di MM viene ricercata nel tratto di strada più vicino al punto P_i . Per riscontrare se il veicolo percorrerà

anche successivamente al tempo i lo stesso tratto di strada, si pongono due condizioni da verificare, che chiameremo Condizioni di Salto (SC): la differenza tra l'azimuth tra due segmenti successivi formati da tre punti GPS maggiore di 45° ; la differenza tra l'*heading* di due punti GPS consecutivi e $\alpha > 45^\circ$. Se sono riscontrate le circostanze si concluderà che il veicolo è in una intersezione tra tratti di strada e si passerà al processo di MM; altrimenti il veicolo sarà ancora sul tratto di strada.

3.1.3 La stima delle coordinate

Lo scopo di questo algoritmo proposto da Quddus *et al.* (2003) è anche quello di determinare la posizione del punto GPS sul tratto di strada candidato e a tale scopo si utilizzano le informazioni dell'angolo θ la velocità v del veicolo, le coordinate della latitudine (*northing*) e della longitudine (*easting*). Dati due punti P_i e P_{i+1} e siano note le coordinate di $P_i(E_i, N_i)$; utilizzando le informazioni di θ e di v del punto P_{i+1} allora si avrà che le sue coordinate sul link saranno:

$$\begin{aligned} E_{i+1} &= E_i + \Delta E_i \\ N_{i+1} &= N_i + \Delta N_i \quad \text{con} \quad \Delta E_i = v \sin \theta \quad \Delta N_i = v \cos \theta \end{aligned}$$

Un altro modo per determinare la posizione è utilizzare le informazioni GPS e proiettarle sul tratto di strada: dato il punto GPS $P_i(e_i, n_i)$ e i nodi del tratto di strada $A(x_i, y_i)$ e $B(x_{i+1}, y_{i+1})$ si proietta perpendicolarmente il punto GPS sul tratto, ricavando il punto $P^{GPS}(e_{GPS}, n_{GPS})$. Utilizzando le coordinate del punto precedente sulla mappa, ovvero $P^{map}=(e_{map}, n_{map})$ e quelle di P^{GPS} si possono stimare le coordinate del punto sulla mappa:

$$\hat{e} = \left(\frac{\sigma_{map,e}^2}{\sigma_{map,e}^2 + \sigma_{gps,e}^2} \right) e_{gps} + \left(\frac{\sigma_{gps,e}^2}{\sigma_{map,e}^2 + \sigma_{gps,e}^2} \right) e_{map}$$

$$\hat{n} = \left(\frac{\sigma_{map,n}^2}{\sigma_{map,n}^2 + \sigma_{gps,n}^2} \right) n_{gps} + \left(\frac{\sigma_{gps,n}^2}{\sigma_{map,n}^2 + \sigma_{gps,n}^2} \right) n_{map}$$

Con $\sigma_{punta\text{coordinata}}^2$ la varianza degli errori associati alla componente longitudine o latitudine dei punti P^{GPS} o P^{map} . Le coordinate di P^{map} si possono ricavare dalla mappa digitale e la velocità del veicolo determinata dal GPS.

In definitiva i passi seguiti per la metodologia di Quddus *et al.* possono così riassumersi:

1. Si cerca il nodo più prossimo al primo punto GPS
2. Se il punto successivo non è un *outlier* si individuano tutti i segmenti di strada che passano per il punto; altrimenti si ipotizza che questo punto sia il punto iniziale e si ritorna al passo 1
3. Si calcola TW e si calcola per quel punto il miglior candidato; si associano i due punti al tratto di strada
4. Si determina la posizione del veicolo sul tratto di strada utilizzando le coordinate stimate (\hat{E}, \hat{N})
5. Si verifica che il punto successivo non sia un *outlier*. Se si verificano le condizioni (SC), si attribuisce il punto allo stesso segmento di strada; si calcolano le coordinate stimate (\hat{E}, \hat{N}) . Se il punto è un *outlier* si ritorna al passo 1 e lo si identifica come il punto iniziale. Tale passo si conclude quando tutti i punti GPS sono riallocati sulla mappa digitale.

3.1.4 La ricerca del percorso più breve

Gli sviluppi dei Sistemi di Trasporto Intelligenti nel campo delle applicazioni dell'ingegneria dei trasporti ha portato la ricerca applicata verso lo studio di algoritmi di ricerca del percorso più breve (*Shortest Path Problem*). Spesso infatti si rende necessaria, in situazioni di scelta in tempo reale, una risposta immediata nell'identificazioni di percorsi più brevi o un ricalcolo veloce del percorso. La velocità della risposta richiede quindi l'utilizzo di algoritmi con un elevata efficienza computazionale, che identifichino nel minor tempo possibile il percorso cercato (appendice B).

Innanzitutto definiamo i termini del problema: sia dato un set di nodi $N=n$ e un numero di archi $A=m$; sia $a=(i,j)$ un arco dal nodo i al nodo j e c_{ij} il costo generico associato a tale arco (pedaggi, lunghezza dell'arco, tempo di percorrenza, ecc). Un percorso dall'origine (O) alla destinazione (d) può essere definito come una sequenza di archi $(O,j), \dots(i,j), \dots(i,d)$ e il tempo di percorrenza dell'intero percorso è dato dalla somma dei tempi di percorrenza dei singoli tratti. Il problema è quello di ricercare il percorso che richiede il minor tempo totale di viaggio dal nodo origine a quello di destinazione. Definiamo anche il costo del percorso dall'origine al nodo i come L_i o L_j se il nodo è identificato con j , con P l'elenco degli archi identificati per il percorso più breve per ogni nodo e con P_i quello per il nodo i . Infine Q è definito come il set di nodi da prendere in considerazione nella ricerca del percorso più veloce.

Molti SPP seguono una procedura standard che si può sintetizzare in quattro passi (Fu *et al.* 2006):

Passo 1 : si pone $i =$ origine; $L_i = 0$; $L_j = \infty \forall j \neq i$; $P_i =$ vuoto; $Q =$ tutti i nodi

Passo 2: si seleziona e si elimina un nodo (i) dal set Q

Passo 3: si analizzano tutti gli archi che partono dal nodo i .

Per ogni arco $a=(i,j)$ se

$$L_i + c_a < L_j \text{ allora}$$

$$L_j = L_i + c_a \text{ e } p_j = a$$

Inserimento del nodo j in Q

Passo 4: se $Q = vuoto$ stop altrimenti ritorna al passo 2.

Questi passi standard tendono però ad avere dei tempi computazionali molto alti in quanto non prendono in considerazione quelle informazioni che possono essere conosciute a priori sulla localizzazione dei punti di origine e destinazione, sulla composizione della struttura della mappa stradale o del percorso. Sono state proposte varie soluzioni di scelta ottimali basandosi su conoscenze aggiuntive del problema da risolvere che possono essere classificate in quattro strategie di ricerca euristica: limitazione dell'area presa in esame, decomposizione del problema di ricerca, limitazione degli archi considerati e alcune combinazioni delle strategie precedenti.

Usando le conoscenze già possedute sugli attributi del percorso più breve dal nodo origine a quello di destinazione, si circoscrive la

ricerca in una zona più limitata o eliminando i nodi intermedi che hanno una bassa probabilità di costituire il percorso più breve, o inserendoli nel set di nodi da prendere in considerazione, ma assegnandogli una bassa probabilità. Il primo metodo è detto del *branch pruning* e limita la zona di indagine ad un'area circolare di forma ellittica, minore della precedente di circa il 20%. Per ridurre l'area di ricerca e migliorare i tempi di computazione e comunque ottenere una soluzione ottimale Fu *et al.* (1996) hanno proposto un criterio in cui la somma L_i del tempo minimo del percorso dall'origine al nodo i sommata al tempo stimato $e_{i,d}$ di percorrenza dal nodo i al nodo di destinazione d non deve superare il limite superiore stimato del tempo minimo di percorrenza dal nodo origine a quello di destinazione $E_{o,d}$:

$$L_i + e_{i,d} \leq E_{o,d}$$

Il secondo passo standard specificato per la ricerca dell'SPP può quindi definire la zona limitata in tale maniera:

passo 2 bis: Selezionare ed eliminare il nodo i con il più basso tempo di percorrenza da Q:

$$\text{se } L_i + e_{i,d} > E_{o,d} \text{ salta al passo 4.}$$

L'accuratezza e l'efficienza di tale selezione dipende delle stime delle funzioni $e_{i,d}$ ed $E_{o,d}$: se la prima è sempre più bassa rispetto al tempo minimo di percorrenza dal nodo i al nodo destinazione e se la seconda è più grande rispetto al tempo minimo di percorrenza tra il nodo origine e quello di destinazione, allora l'algoritmo è ottimale. La definizione di queste due funzioni

diventa quindi fondamentale: gli autori propongono di definire le due funzioni come segue:

$$e_{i,d} = \frac{D_{i,d}}{V} ; \quad E_{o,d} = Ke_{o,d}$$

con V uguale alla velocità media di percorrenza, D la distanza in termini euclidei tra il nodo i e il nodo di destinazione d e K un parametro controllabile e definibile dall'esterno.

Un problema emerso nell'applicazione del *branch pruning* è la circostanza che l'algoritmo potrebbe arrestarsi senza aver definito una soluzione; una soluzione proposta è quella di incrementare il fattore K e iniziare nuovamente la ricerca dal nodo origine, oppure di inserire nella lista Q dei nodi da analizzare quelli che sono stati eliminati ad ogni iterazione:

$$K_{n+1} = \frac{\max_{i \in P_n} \{L_i + e_{i,d}\}}{e_{o,d}}$$

dove P_n è definito come l'insieme dei nodi eliminati nell'iterazione n . Bisogna tener presente però che maggiore è il fattore K maggiore sarà l'area di analisi e migliore la soluzione ottimale ma anche il tempo computazionale, che è un parametro fondamentale da tener sotto controllo nello studio in tempo reale dei problemi ITS.

L'algoritmo A^* è un differente metodo di ricerca del percorso più breve limitando l'area di ricerca. (Hart, Nilsson, 1968); viene utilizzata la funzione di valutazione euristica $F_i = L_i + e_{i,d}$ con L_i il tempo di spostamento sull'arco i dall'origine al nodo i che definisce per il nodo i quanto sia un buon candidato per il più breve

percorso: più basso è tale valore maggiore sarà la sua probabilità di far parte del percorso. Anche per l'algoritmo A^* si può applicare un percorso standard variando il passo 3 in tal modo:

Passo 3bis: analisi di tutti gli archi uscenti dal nodo i ; per ogni arco i,j

$$\text{Se } L_i + c_{i,j} + e_{j,d} < F_j$$

$$L_j = L_i + c_{i,j}; \quad P_j = L_i + c_{i,j} + e_{i,d}; \quad P_j = a$$

Il nodo j viene inserito in Q

La accuratezza del metodo dipende dalla funzione euristica F_i e quindi dalla funzione stimata e ; se ipotizziamo di considerare $e_{i,d} = D_{i,d}/V$ con V il parametro che può subire variazioni, in quanto la distanza è calcolata come distanza euclidea tra coordinate, allora maggiore sarà la velocità media maggiore sarà l'area di ricerca e il tempo computazionale.

L'area presa in considerazione di forma anch'essa ellittica per la ricerca del percorso più breve sarà definita come:

$$L_i + e_{i,d} \leq L_d$$

Confrontando il set di nodi presi in considerazione dall'algoritmo A^* e dal *branch pruning* si osserva che, se il primo prende in considerazione tutti i nodi presenti sulla frontiera dell'ellisse, il secondo ne elimina alcuni, in funzione della funzione che rappresenta la zona di analisi e quindi risulta essere più veloce.

Se la distanza tra origine e destinazione è notevole o se i nodi presenti sono un numero considerevole, i tempi computazionali crescono velocemente. Per questo motivo è possibile decomporre l'analisi generale in più sottoproblemi. Tra le varie strategie di decomposizione sono state applicate metodi di ricerca bi-direzionali e di sotto obiettivo (*subgoal methods*).

La procedura bi-direzionale prevede la suddivisione in due problemi separati: il primo parte dal nodo origine mentre il secondo parte dal nodo destinazione; si esaminano contemporaneamente tutti i percorsi che partono dall'origine e terminano nel punto destinazione e i punti adiacenti connessi. Il processo si ripete finché un percorso che parte dall'origine non incontra un altro che è partito dal nodo destinazione e quando le due procedure si incontrano allo stesso stadio, la soluzione ottimale è raggiunta (Nicholson, 1966).

Nella ricerca del percorso più breve tutti gli archi che si diramano dai nodi vengono presi in considerazione ed analizzati per il contributo che possono dare al percorso ottimale. Un diverso approccio prevede che non si prendano in considerazione quegli archi che hanno una bassa probabilità di far parte del percorso ottimale o che non possono essere usati in pratica (ad esempi tratti di strada senza uscita). Lo sviluppo di questo approccio è stato possibile mutuando la strategia di ricerca gerarchica utilizzata nel campo dell'intelligenza artificiale, ovvero l'idea che per risolvere un problema complesso si analizzano i particolari essenziali del problema stesso, senza tener conto di livelli di dettaglio molto particolari e successivamente completare l'analisi con un livello maggiore di dettaglio. La metodologia è alquanto complessa se applicata alla ricerca del percorso più breve per la difficoltà di

modellizzare con tale tecnica un reale mappa stradale e per controllare la ricerca quando si passa da un livello di dettaglio ad un altro.

Diversi autori hanno proposto dei criteri per creare i diversi livelli di dettaglio per le mappe stradali: Chou *et al.*(1998) hanno definito i diversi livelli in base alla lunghezza degli archi: i tratti di strada lunghi formano il livello di dettaglio alto e i tratti meno lunghi formano un livello basso di dettaglio; Liu (1997) e Jagadeesh *et al.* (2002) raggruppano invece tutte le strade in due livelli rispetto ai loro attributi (limiti di velocità, numero di corsie, ecc.).

3.2 L'approccio probabilistico

L'approccio topologico non sempre è il metodo più adatto per la procedura di MM, soprattutto se la rete stradale è complessa o quando sono presenti degli outliers. L'approccio probabilistico è risultato un metodo spesso adottato per migliorare sia l'individuazione sulla rete del punto iniziale del percorso sia dei successivi. Con tale metodologia si crea una zona in cui si prevede con una certa probabilità che cadano tutti quei tratti di strada, possibili candidati del matching con il punto GPS che si considera. In tal modo si prendono in considerazione non solo tutti i punti GPS e la topologia di rete, ma anche gli errori di misura che possono provocare una errata attribuzione di un punto GPS ad un nodo o un tratto di strada. Sono state avanzate e sperimentate diverse proposte, una di queste è di definire una regione di confidenza intorno a tutti o ad alcuni particolari punti GPS; la grandezza della regione è funzione della varianza degli errori: tutti i tratti di strada che ricadono all'interno della regione sono valutati come possibili candidati, utilizzando come parametri

di riferimento della loro bontà la distanza, l'*heading*, il tratto precedente o la velocità.

La creazione di intervalli di confidenza può essere visto come un approccio di base di tale metodologia; i parametri con cui vengono valutati i diversi tratti di strada candidate possono variare in base alle informazioni che si hanno a disposizione, alle caratteristiche del problema o ai parametri che si ritiene possano meglio evidenziare le caratteristiche dei tratti di strada. Generalmente il problema del MM si suddivide in due momenti: il primo che risolve l'identificazione del punto iniziale del percorso e il secondo che, sulla base dell'informazione del punto iniziale, analizza la sequenza degli ulteriori punti GPS e definisce il percorso sulla mappa digitale. Il primo momento è necessario in quanto la corretta identificazione del primo punto è fondamentale per il corretto prosieguo del MM, ma la sua identificazione è problematica in quanto non è possibile confrontare il primo punto GPS con nessun altro punto precedente o con informazioni storiche acquisite: una scorretta attribuzione del primo punto si può ripercuotere inevitabilmente su tutto il percorso.

Pyo *et al.* (2001) introducono la tecnica dell'ipotesi multipla, secondo cui mediante l'identificazione di una regione di errore sono considerate diverse strade candidate al MM per il punto considerato, e ad ognuna viene attribuito un punteggio. Per limitare il numero di strade candidate Pyo *et al.* non definiscono alcun criterio o valore di soglia (*threshold*) che permetta di diminuire il numero di strade candidate, ma basano la diminuzione delle strade sul punteggio attribuito al miglior candidato. Con tale criterio il numero di strade candidate risulta essere molto alto circostanza che, insieme al lavoro computazionale e al tempo impiegato, rendono l'applicazione estremamente costosa in termini di iterazioni e di tempo. Per rendere più veloce l'algoritmo del MM

utilizzando sempre la regione di confidenza, Quddus *et al.* (2006) hanno proposto un algoritmo che si basa sulla creazione di una regione ellittica dell'errore solo per quei punti GPS che si trovano in corrispondenza di incroci o per il punto iniziale. Per ridurre al minimo gli errori soprattutto negli incroci o in svolte del veicolo, hanno sviluppato dei criteri per la loro identificazione. Inoltre propongono questa metodologia anche per quei percorsi GPS in cui si registrano basse velocità di spostamento come nel caso di reti stradali complessi o di grandi centri urbani.

3.2.1 Un esempio di algoritmo probabilistico

Quddus *et al.* propongono un algoritmo probabilistico avanzato laddove il veicolo percorra il suo tragitto a bassa velocità e di conseguenza l'*heading* del GPS dia valori poco affidabili. Il veicolo è dotato di GPS DR e l'integrazione avviene mediante un Filtro Kalman Esteso, che prende come valori di input sia le informazioni del GPS che del DR e gli errori associati e definisce in output le coordinate del punto, la velocità del veicolo, l'*heading* e la loro varianza degli errori. L'algoritmo prende in considerazione alcune le informazioni derivate sia dalla mappa digitale che dal GPS, come la topologia di rete, gli archi e i nodi della rete e la varianza dell'errore della mappa. L'algoritmo si sviluppa in due stadi: il primo identifica, mediante la creazione di regioni di errore, il punto iniziale o il percorso sulla mappa quando il veicolo attraversa gli incroci o genera una manovra di svolta; il secondo individua la strada utilizzando gli ulteriori punti GPS.

Nel primo stadio viene analizzato il primo punto del percorso, in quanto la sua scorretta determinazione può comportare ripercussioni su tutto il MM successivo. Per ottenere ciò, si definisce una regione di confidenza intorno al punto considerato, all'interno della quale tutti i tratti

di strada saranno visti come possibili candidati per l'associazione del punto GPS al tratto di strada. Se la regione definita non contiene nessun tratto di strada si concluderà che il veicolo è fuori da ogni possibile strada conosciuta. Se invece ci sono più tratti di strada, si calcola per ogni tratto la differenza tra l'*heading* del veicolo e quello del tratto di strada: se la differenza è minore di 45° allora il tratto viene considerato un candidato; si formerà quindi un set di possibili candidati. Per definire quello più appropriato si fa uso delle misure di connettività della mappa, la distanza minore con il punto, le informazioni storiche. Queste ultime sono conosciute solo per quei punti analizzati dall'algoritmo che non rientrano nella categoria di primo punto, ovvero quando si è in presenza di un incrocio o di una manovra in un incrocio. Se nella regione di confidenza ricade unicamente un tratto di strada si opera un test di connettività della mappa che, se non confermato, produce un incremento della regione di confidenza e l'applicazione nuovamente dell'algoritmo iniziale.

Il secondo algoritmo viene applicato se il veicolo non si trova in un incrocio o se non ci sono cambiamenti di direzione; inoltre se, la distanza dall'incrocio è maggiore di 10 metri e la misura angolare misurata dal giroscopio è minore di 15° . Il momento fondamentale è legato all'individuazione del veicolo negli incroci o durante una manovra di svolta che può essere rilevato utilizzando l'informazione sulla velocità del veicolo. Ricordando che se la velocità è bassa l'*heading* del veicolo non è più affidabile, bisogna trovare un valore di soglia per definire la velocità minima al di sotto della quale l'*heading* del veicolo ottenuto dal GPS non è più utilizzabile.

Taylor (2001) sperimenta che la correlazione tra i due *heading* è bassa quando il veicolo viaggia al più a 8 km/h; Quddus pone invece come valore di soglia la velocità di 10.8 km/h, valore frutto di diversi

test svolti su diverse tipologie di strade. Un primo test è stato realizzato utilizzando la deviazione assoluta dell'*heading* del veicolo, calcolata come differenza tra l'*heading* del veicolo ottenuto grazie alla mappa digitale e quello derivante dalla mappa digitale. I risultati sperimentali hanno mostrato che questa differenza è sempre minore di 30° quando il veicolo viaggia ad una velocità superiore a 10.8 km/h. Un'ulteriore prova è stata compiuta per conoscere come cambia l'*heading* su tratti rettilinei, per esempio in caso di sorpasso: non si è constatato una variazione significativa nel trend dell'*heading*. Infine, osservazioni sperimentali su un veicolo in movimento in un incrocio o in manovra di svolta hanno documentato che il tempo necessario per completare una manovra quando si è in presenza di un incrocio a quattro vie e il veicolo transita ad una velocità superiore alla velocità minima, è di 4 secondi. Questo tempo è funzione, però del tipo di incrocio e della velocità di spostamento: in generale si stima il tempo di percorrenza tra i due e i cinque secondi.

Come accennato in precedenza per determinare la posizione del veicolo sulla mappa digitale si può proiettare perpendicolarmente il punto GPS $P_i(e_i, n_i)$ sul tratto di strada con nodo iniziale $A(x_1, y_1)$ e nodo finale $B(x_1, y_2)$, ricavando il punto $P^{GPS}(e_{GPS}, n_{GPS})$. Utilizzando le coordinate del punto precedente sulla mappa, ovvero $P^{map}=(e_{map}, n_{map})$ (mediante l'ultimo nodo e la velocità del veicolo rilevata dal GPS) si possono stimare le coordinate del punto sulla mappa. La Figura 18 evidenzia i punti utilizzati.

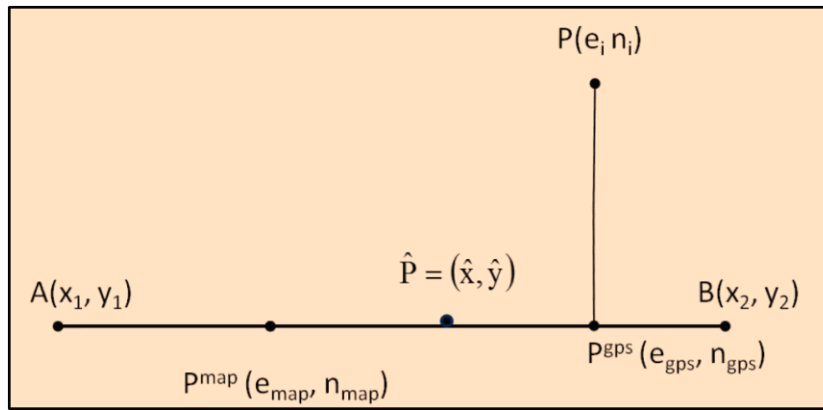


Figura 18: Punti per la stima delle coordinate

Sia P^{map} che P^{GPS} sono due punti che rappresentano la posizione del veicolo sulla mappa digitale, ma non la vera posizione, così come qualsiasi altro punto associato al tratto di strada rappresenterà comunque la stima del punto GPS. Le coordinate *easting* di entrambi i punti possono essere espresse come:

$$e_{map} = e + \omega_{map,e}$$

$$e_{gps} = e + \omega_{gps,e}$$

con $\omega_{map,e}$ l'errore dell'*easting* associato alla mappa digitale e alla velocità e $\omega_{GPS,e}$ l'errore relativo alla componente *easting* associato alla proiezione del punto GPS. Il valore medio di questi due errori è dato dal prodotto del coefficiente di correlazione tra gli errori e le deviazioni standard dei rispettivi punti, ovvero:

$$E(\omega_{map,e}, \omega_{gps,e}) = \rho \sigma_{map,e} \sigma_{gps,e}$$

Il valore di e può essere solo stimato mediante una stima ottimale che viene definita mediante una funzione lineare:

$$\hat{e} = k_1 e_{map} + k_2 e_{gps}$$

I parametri k_1 e k_2 devono essere ricavati e sono indipendenti da e . La differenza tra la stima ottimale e il valore di e definirà l'errore stimato:

$$\tilde{e} = \hat{e} - e$$

La stima di \tilde{e} deve essere non distorta ovvero:

$$E[\tilde{e}] = E[k_1(e + \omega_{map,e}) + k_2(e + \omega_{gps,e}) - e] = 0$$

ovvero quando $k_1 = 1 - k_2$ se $E(\omega_{map,e}) = E(\omega_{gps,e}) = 0$ e $E(e) = 0$.

Il criterio di ottimalità è quello di minimizzare $E(\tilde{e}^2)$. Sostituendo il vincolo di non distorsione, azzerando la derivata prima si avrà:

$$k_1 = \frac{\sigma_{gps,e}^2 - \rho \omega_{map,e} \sigma_{gps,e}}{\sigma_{map,e}^2 + \sigma_{gps,e}^2 - 2\rho \omega_{map,e} \sigma_{gps,e}} \quad e \quad k_2 = \frac{\sigma_{map,e}^2 - \rho \omega_{map,e} \sigma_{gps,e}}{\sigma_{map,e}^2 + \sigma_{gps,e}^2 - 2\rho \omega_{map,e} \sigma_{gps,e}}$$

che sostituiti nella stima ottimale definiscono la coordinata *easting* di \hat{P} .
Con lo stesso procedimento si ricavano la coordinata *northing*.

L'errore quadratico medio della stima, ovvero la varianza di \tilde{e} , sarà quindi:

$$E(\tilde{e}^2) = \sigma_e^2 = \frac{\sigma_{map,e}^2 - \sigma_{gps,e}^2 (1 - \rho)}{\sigma_{map,e}^2 + \sigma_{gps,e}^2 - 2\rho \omega_{map,e} \sigma_{gps,e}}$$

Se non esiste correlazione tra $\omega_{map,e}$ e $\omega_{gps,e}$ e $\rho=0$ allora i valori di k_1 e k_2 saranno:

$$k_1 = \left(\frac{\sigma_{gps,e}^2}{\sigma_{map,e}^2 + \sigma_{gps,e}^2} \right) \quad k_2 = \left(\frac{\sigma_{map,e}^2}{\sigma_{map,e}^2 + \sigma_{gps,e}^2} \right)$$

Se $\sigma_{map,e} = \sigma_{gps,e}$ la stima ottimale sarà :

$$\hat{e} = \frac{e_{map} + e_{gps}}{2}$$

3.2.2 La regione dell' errore

Per calcolare l'ampiezza della regione di errore si considerano le coordinate x e y del punto intorno al quale definirla: la direzione dell'orientamento dell'ellisse è definita dall'angolo β e quindi dalla direzione del piano di riferimento (Nord) e dalla direzione del veicolo: se le coordinate x e y non sono correlate, allora gli assi dell'ellisse saranno paralleli al sistema di riferimento del veicolo. La Figura 19 mostra una regione di confidenza il cui orientamento è nella direzione del veicolo ed ha una inclinazione pari a β rispetto al *North* del sistema assoluto di riferimento.

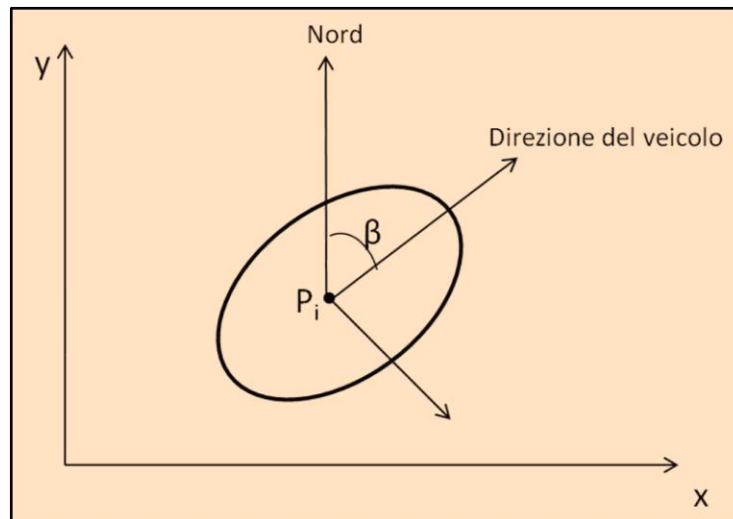


Figura 19: Direzione di una regione di errore

Dati σ_x^2 e σ_y^2 le varianze degli errori delle coordinate di P_i e σ_{xy} , σ_{yx} le covarianze, consideriamo la matrice delle varianze-covarianze:

$$A = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

Il polinomio caratteristico di A è dato da:

$$\begin{aligned} D(\lambda) &= \text{Det}(A - \lambda I) = \begin{vmatrix} \sigma_x^2 - \lambda & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 - \lambda \end{vmatrix} = (\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - \sigma_{xy}^2 = 0 \end{aligned}$$

Gli autovettori della matrice A sono le radici dell'equazione caratteristica A , ovvero:

$$\begin{aligned} \lambda &= \frac{(\sigma_x^2 + \sigma_y^2) \pm \sqrt{(\sigma_x^2 + \sigma_y^2)^2 - 4(\sigma_x^2 \sigma_y^2 + \sigma_{xy}^2)}}{2} \text{ da cui} \\ a &= \sqrt{\frac{1}{2}(\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2})} \\ b &= \sqrt{\frac{1}{2}(\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2})} \end{aligned}$$

con a e b le radici quadrate di λ che rappresentano il semiasse maggiore e minore rispettivamente dell'ellisse. L'angolo β può essere determinato in tal modo:

$$\beta = \frac{\pi}{2} - \frac{1}{2} \arctan\left(\frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}\right)$$

Per compensare gli errori dovuti alla ricezione del segnale GPS, a e b devono essere moltiplicati per un fattore cosiddetto di espansione $\hat{\sigma}_o$. Allora se il fattore di compensazione si distribuirà normalmente, l'errore

standard dell'ellisse corrisponderà al 39% della regione di confidenza. Zhao (1997) dimostra che per ottenere un livello di confidenza pari al 99% il fattore di compensazione dovrà essere pari a 3,03.

Ma gli errori che si possono verificare nell'individuazione del percorso GPS di un veicolo possono derivare anche da altri fattori quali errori della mappa digitale, di imputazione dei dati o nei dati stessi. Per tale motivo è necessario moltiplicare per ulteriori fattori di compensazione. Il problema può essere aggirato espandendo la regione di confidenza e rendendola rettangolare; in particolare si costruisce un rettangolo intorno all'ellisse così definita, con le misure pari a quelle dei semiassi dell'ellisse stessa.

Come abbiamo accennato in precedenza, è possibile che nell'ellisse dell'errore definita intorno ad un punto GPS con particolari caratteristiche (punto iniziale o rilevazione in un incrocio o durante una svolta) possano verificarsi le seguenti situazioni (Figura 20 Figura 21):

- Nessun tratto di strada cade nell'ellisse

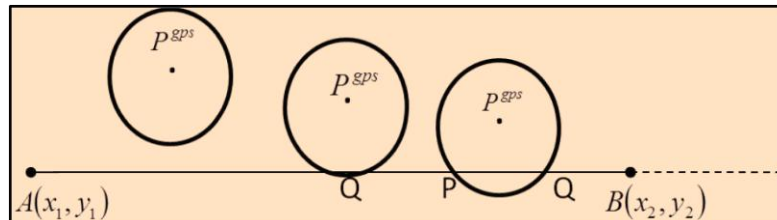


Figura 20: La regione dell'errore e il tratto di strada

- L'ellisse è tangente al tratto di strada
- L'ellisse è secante al tratto di strada
- L'ellisse interseca la strada in un punto reale e in un punto sul suo prolungamento (punto ipotetico)
- L'ellisse è tangente al prolungamento del tratto di strada
- Il prolungamento del tratto di strada è secante all'ellisse

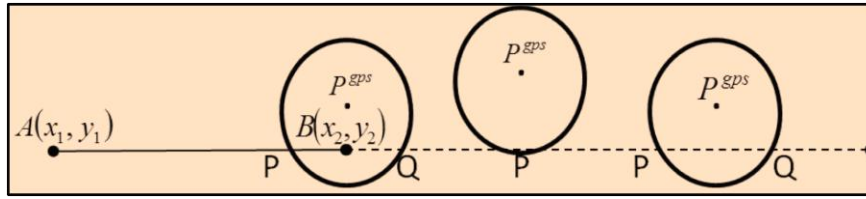


Figura 21: La regione dell'errore e il tratto di strada

Per verificare la presenza di una delle situazioni descritte definiamo con $A(x_1, y_1)$ e $B(x_2, y_2)$ il nodo iniziale e finale di un tratto di strada, $P^{GPS}(e_{GPS}, n_{GPS})$ il punto GPS, con $\Delta x = x_1 - x_2$ e $\Delta y = y_1 - y_2$. Per semplicità definiamo la regione di errore come una regione circolare (Figura 20 e 21).

Allora l'equazione del cerchio con centro in P e l'equazione della retta passante per A e B saranno:

$$(x - e_{gps})^2 + (y - n_{gps})^2 = r^2$$

$$y = mx + c \quad \text{con} \quad m = \frac{\Delta x}{\Delta y} \quad e \quad c = \frac{x_1 y_2 - y_1 x_2}{\Delta x}$$

I punti di intersezione tra il cerchio e la retta saranno:

$$a = \frac{m^2 n_{gps} + e_{gps} - mc}{1 + m^2} \pm \frac{\sqrt{r^2 - g^2}}{\sqrt{1 + m^2}}$$

$$b = \frac{m^2 n_{gps} + m e_{gps} + c}{1 + m^2} \pm \frac{m \sqrt{r^2 - g^2}}{\sqrt{1 + m^2}}$$

Con $g = \frac{n_{gps} - m e_{gps} - c}{\sqrt{1 + m^2}}$. Si potranno determinare i seguenti casi:

- $(r^2 - g^2) < 0$ non c'è nessuna intersezione tra il tratto di strada e la regione di errore
- $(r^2 - g^2) = 0$ esiste un punto di tangenza

- $(r^2 - g^2) > 0$ il tratto sarà secante alla regione di errore

Se si verificano le seguenti condizioni

$$|x_1| \leq a \leq |x_2| \text{ o } |x_2| \leq a \leq |x_1|$$

$$|y_1| \leq a \leq |y_2| \text{ o } |y_2| \leq a \leq |y_1|$$

Allora il tratto di strada della mappa digitale intersecherà realmente la regione dell'errore.

3.3 Applicazioni del modello *fuzzy* al *map matching*

Diversi autori (Zhao,1997; Kim *et al.* 1998; Kim e Kim,2001; Syed e Cannon, 2004; Quddus *et al.*, 2006; Zhang & Gao, 2008) hanno utilizzato la logica *fuzzy* e i modelli *fuzzy* per affrontare il problema della corretta identificazione di un percorso GPS ad una mappa digitale. In questo paragrafo si confrontano tre algoritmi, il primo proposto da Quddus *et al.* (2006), il secondo da Syed e Cannon (2004) e il terzo da Zhang e Gao (2008).

Quddus *et al.* suddividono la procedura in due stadi: il primo in cui si identifica il punto iniziale del percorso sulla mappa e il secondo che individua il percorso successivo. Per identificare correttamente il primo punto propongono l'algoritmo *Initial Map matching Process* (IMP) e definiscono una regione ellittica o rettangolare intorno al punto: se nella regione non è contenuto alcun tratto di strada allora il veicolo starà percorrendo strade non presenti sulla mappa digitale; se invece sono presenti più tratti di strada inizia il processo di identificazione del corretto

abbinamento mediante un modello *fuzzy*. Le variabili input e i relativi qualificatori sono specificati nella Tabella 15:

Tabella 15: Variabili fuzzy dell'IMP

Variabili	Simbolo	Qualificatori
Velocità del veicolo	v	<i>Nulla</i> <i>Bassa</i> <i>Alta</i>
Errore dell' <i>heading</i>	HE	<i>Piccolo</i> <i>Grande</i>
Distanza perpendicolare	PD	<i>Piccola</i> <i>Grande</i>
HDOP	HDOP	<i>Buona</i> <i>Non Buona</i>

Il modello applicato è quello di Sogeno di ordine zero e quindi definisce una funzione di output *L1* che può assumere tre costanti il cui significato è basso, medio e alto. Le regole che seguono sono quelle che definiscono il modello:

If [v è alto]and [HE è piccolo]THEN [L1 è medio]

If [v è alto]and [HE è grande]THEN [L1 è basso]

If [HDOP è buono]and [PD è piccolo]THEN [L1 è medio]

If [HDOP è buono]and [PD è grande]THEN [L1 è basso]

If [HE è piccolo]and [PD è piccolo]THEN [L1 è alto]

If [HE è grande]and [PD è grande]THEN [L1 è basso]

Viene utilizzato il metodo del “minimo” per calcolare il grado di applicabilità di ogni regola *fuzzy*; per ottenere il valore *crisp* dell’output viene usato il metodo della media pesata: quel valore che sarà maggiore di tutti rappresenterà il corretto MM tra tutti i tratti di strada candidati. Per diminuire quanto più possibile l’errore che si potrebbe commettere nell’identificazione del primo punto, gli autori propongono di analizzare alcuni punti GPS successivi al primo e, mediante l’applicazione del modello *fuzzy*, verificare se tutti si associano allo stesso tratto di strada. Se ciò avviene allora il primo punto sarà fissato sul quel tratto di strada e si passerà all’attuazione del secondo algoritmo per l’identificazione del percorso, mediante i punti GPS successivi al primo.

Questo secondo algoritmo, detto *Subsequent Map matching Process* (SMP), si suddivide in due segmenti: il primo viene applicato se i punti possono essere associati al tratto di strada precedente, mentre il secondo se si è in presenza di incroci. Il primo tipo di SMP è funzione di 4 parametri: la direzione del veicolo (θ), la variazione di giroscopio ($\Delta\theta$), la distanza sul tratto di strada tra la posizione dell’ultima corrispondenza e l’incrocio ($d1$) e la velocità del veicolo (v). La Figura 22 evidenzia le misure considerate.

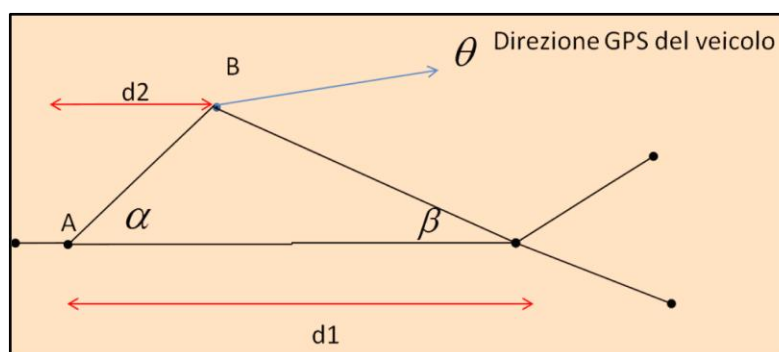


Figura 22: Variabili considerate per l'SMP

La differenza tra $d1$ e $d2$ serve ad evidenziare se il veicolo sta attraversando un incrocio. Infatti se la differenza è negativa si può ipotizzare che il veicolo abbia superato un incrocio; stessa circostanza se gli angoli α e β sono minori di 90° . In definitiva si definiscono sette variabili *fuzzy*: la velocità (v), l'incremento dell'*heading* (HI), il tasso del giroscopio ($\Delta\theta$), la differenza tra le distanze (Δd), i valori degli angoli e l'HDOP. L'output è definito con $L2$ e viene utilizzato, come per il IMP, il modello di Sogeno di ordine zero che può assumere tre diverse costanti: $Z1$, $Z2$, $Z3$. Le regole *fuzzy* sono le seguenti:

If [$\Delta\theta$ è piccolo] *and* [α è inferiore a 90°] *and* [β è inferiore a 90°] *THEN*
[$L2$ è alto]

If [$\Delta\theta$ è piccolo] *and* [Δd è positivo] *and* [α è inferiore a 90°] *THEN* [$L2$ è basso]

If [$\Delta\theta$ è piccolo] *and* [Δd è positivo] *and* [β è inferiore a 90°] *THEN* [$L2$ è basso]

If [HI è piccolo] *and* [α è inferiore a 90°] *and* [β è inferiore a 90°] *THEN*
[$L2$ è alto]

If [HI è piccolo] *and* [Δd è positivo] *and* [α è inferiore a 90°] *THEN* [$L2$ è basso]

If [HI è piccolo] *and* [Δd è positivo] *and* [β è inferiore a 90°] *THEN* [$L2$ è basso]

If [$\Delta\theta$ è alto] *and* [α è inferiore a 90°] *and* [β è inferiore a 90°] *THEN* [$L2$ è basso]

If [HI è grande] *and* [α è inferiore a 90°] *and* [β è inferiore a 90°] *THEN*
[$L2$ è basso]

If [HDOP è buono] *and* [v è zero] *THEN* [L2 è alto]

If [HDOP è buono] *and* [Δd è negativo] *and* *THEN* [L2 è medio]

If [HDOP è buono] *and* [Δd è positivo] *and* *THEN* [L2 è basso]

If [v è alta] *and* [HI è piccola] *THEN* [L2 è medio]

If [HDOP è buono] *and* [v è alta] *and* [HI è 180°] *and* [$\Delta\theta$ è alto] *THEN*
[L2 è basso]

Quddus definisce un valore soglia derivato empiricamente, per definire se il punto GPS può essere associato al tratto di strada precedente. Se il valore $L2$ è più alto rispetto al valore di *threshold* allora il punto GPS verrà associato al tratto di strada identificato.

Il secondo tipo di algoritmo SMP viene utilizzato se il veicolo è presente prima o in un incrocio. Oltre ai parametri definiti nel IMP si aggiungono altre due variabili: la informazioni di connettività e l'errore di distanza. Le prime prendono in considerazione se i tratti candidati sono direttamente connessi al tratto precedente o altre informazioni relative alle caratteristiche della strada, mentre il secondo prende in considerazione la differenza della distanze tra il punto precedente già connesso alla strada e la proiezione perpendicolare del punto GPS considerato sui diversi archi di strada possibili candidati. L'output del modello di Sogeno di ordine zero può assumere tre valori $L3$, definiti come basso, medio e alto. Il tratto di strada con il valore $L3$ più alto è il candidato migliore. Le regole *fuzzy* aggiunte a quelle già presenti nel IMP sono:

If [la connettività è bassa] *THEN* [L3 è basso]

If [la connettività è alta] THEN [L3 è alto]

If [errore di distanza è basso] THEN [L3 è alto]

If [errore di distanza è alto] THEN [L3 è basso]

Tutto il processo viene implementato usando il toolbox di Matlab per l'applicazione dei modelli *fuzzy*, utilizzando come MF l'*Adaptive Neuro Inference System* (ANFIS). Quddus *et al.*(2006) applicano questa metodologia ad un caso reale e valutano che la corretta identificazione del percorso alla mappa digitale sia del 99.2% .

3.3.1 Un modello fuzzy per dati affetti da “effetto canyon”

Syed e Cannon (2004) propongono un algoritmo avanzato utilizzando anche loro il modello *fuzzy* laddove il percorso sia disturbato dall'effetto canyon, di difficile modellizzazione. Definiscono due algoritmi: il primo per la ricerca del punto iniziale e il secondo per la ricerca del percorso successivo. Si definisce una regione in cui ricadono tutti i tratti di strada nel raggio di 50 metri e questi saranno i possibili candidati. Vengono presi in considerazione quattro tipi di variabili: la distanza tra il punto e i diversi tratti di strada (DE); la velocità e la direzione rispetto ai tratti di strada (VD), il cambiamento dell'*heading* (HC) ottenuto dal giroscopio, il tempo (T), ottenuto dal GPS. Il *fuzzy set* è composto da:1) la variazione di *heading* nominale (NH), 2) l'appartenenza al set di strade prese come riferimento (B), 3) la similarità della velocità e della direzione all'orientamento del tratto di strada (S), 4) l'alto numero dei periodi temporali (T).

Le regole che gli autori definiscono sono le seguenti:

If [HC è nominal] AND[S è per un tempo T considerevole vera] AND [S é vera] THEN

*[Z=MF1*MF2*MF3]*

Con *MF1*, *MF2* e *MF3* gli output dei *fuzzy sets* 1), 2) e 4) rispettivamente.

La seconda regola viene presa in considerazione se l'output è al di sopra di un valore di soglia empiricamente derivato di 0.7, altrimenti questo passo è ripetuto finché il valore *LI* non superi il valore di soglia. La seconda regola è la seguente:

If [HC è alto] AND[S è vera] AND [LI futuri è più alto dell'attuale] THEN [Z= è il primo tratto]

Dopo aver individuato il corretto tratto di strada da associare al primo punto GPS, il secondo sub-algoritmo permette di trovare il restante percorso; esso utilizza le stesse variabili (*fuzzy inputs*) e definisce il *fuzzy sets* composto da: 1) la prossimità tratto di strada (*P*); 2) un piccolo cambiamento dell'*heading* (*HC*), 3) la distanza media percorsa sul tratto di strada (*AD*), 4) la distanza maggiore percorsa sul tratto di strada (*LD*), 5) la differenza tra la variazione dell'*heading* e l'angolo formato tra l'attuale e il successivo tratto di strada (*HD*). Le regole sono:

If [AD è medio] AND[P è alto] AND [HD è basso] THEN [

$$Z = MF1 \cdot \left(\frac{1.5}{1 + \frac{1}{1 + MF2}} \right) \cdot MF3]$$

con *MF1*, *MF2* e *MF3* gli output rispettivamente dei *fuzzy sets* 3), 1) e 2) del secondo sub-algoritmo.

La seconda regola è:

If [LD è vero] AND [HD è piccolo] THEN [Z =MF1·MF2]

Con MF1 e MF2 gli output dei fuzzy set 4) e 5) del secondo sub-algoritmo.

Gli autori utilizzano il toolbox di Matlab per l'applicazione dei modelli *fuzzy*, e come MF l'ANFIS, che utilizza tecniche come i minimi quadrati e l'algoritmo del *back propagation* per la ricerca dell'ottimo. Gli autori sottolineano che le condizioni di inizializzazione del GPS definiscono il tempo di posizionamento del primo punto sulla mappa digitale, che la variazione dell'*heading* identifica un cambio di strada e che indirettamente aiutano ad identificare il segmento corretto. Inoltre se ci sono perdite di segnale GPS o degli *outliers* persistenti, l'algoritmo riprende dal primo sub-algoritmo. Comunque questa metodologia presente dei risultati poco soddisfacenti se l'angolo tra i diversi tratti di strada è molto piccolo, in quanto la variazione potrebbe non essere percepita dai giroscopi comuni.

3.3.2 Una variante al modello fuzzy logic

Il terzo algoritmo preso in considerazione è quello proposto da Zhang e Gao (2008) che utilizzano la posizione GPS, l'*heading* e la velocità per determinare la posizione del veicolo; inoltre viene utilizzata anche l'informazione storica della traiettoria del veicolo. Le informazioni topologiche sono usate per identificare il percorso negli incroci. Come

negli altri algoritmi il problema viene suddiviso in due passi: l'identificazione del primo tratto di strada e poi quelli successivi. Il *fuzzy set* preso in considerazione è composto da: la velocità (v), l'errore dell'*heading* (HE), la distanza perpendicolare (PD) e l'HDOP (HDOP). La tabella riassuntiva per l'IMP di Quddus riassume i qualificatori anche per questo algoritmo. Le regole sono le stesse dell'algoritmo IMP di Quddus *et al.* così come le costanti di output. Il secondo sub-algoritmo prende in considerazione la velocità del veicolo, l'incremento dell'*heading* (HI), le informazioni del giroscopio (G), che assume i valori basso e alto, e l'HDOP.

If [v è alto]*and* [HI è piccolo]*THEN* [L1 è medio]

If [v è alto]*and* [HE =180°] *and* [HDOP è buono] *and* [GR è alto]*THEN* [L1 è alto]

If [$v=0$]*and* [HDOP è buono]*THEN* [L1 è alto]

If [v è alto]*and* [GR è piccolo]*THEN* [L1 è medio]

If [$v=0$]*and* [GR è piccolo]*THEN* [L1 è alto]

Per identificare il percorso reale quando la velocità è bassa, come negli incroci, gli autori definiscono una regione ellittica intorno al punto GPS utilizzando le sue coordinate e la formulazione di Quddus (2006); calcolano la distanza tra il punto GPS e le possibili strade, ma non la distanza più piccola, così da ridurre la complessità operativa.

4. Una proposta operativa: la penalizzazione e il *windowing*

In questo capitolo viene descritta una proposta metodologica al problema e si mostrano i primi risultati su dati sperimentali: la procedura si basa sul confronto tra un percorso GPS nella sua totalità e un set di strade candidate che ricadono in una finestra definita intorno al percorso stesso. Grazie ai singoli punti rilevati, si definisce una spezzata che rappresenta l'andamento del veicolo; intorno a tale percorso si definisce un'area di osservazione o finestra, che sarà funzione anche di un parametro definito a priori. Questa finestra conterrà il percorso GPS e una serie di strade della mappa digitale che rappresenteranno i candidati al *map matching*: le aree tra il percorso e ogni singola strada vengono interpretate come un costo che si sostiene per ritornare su un tracciato stradale definito; quella strada che renderà minimo tale costo rappresenterà la strada migliore. L'approccio proposto tende a coniugare sia la gestione del problema dal punto di vista geometrico senza tralasciare la storia del percorso e il suo obiettivo finale. Il percorso viene ricostruito sulla mappa digitale basandosi anche su fattori oggettivi e comportamentali.

4.1 Il costo dello scostamento

Si ipotizzi di avere una sequenza di punti GPS rilevati e associati ad una mappa digitale. Questa sequenza potrà essere composta da una parte di rilevazioni che coincidono perfettamente con gli archi orientati di una o più strade della mappa e altri punti che, per le valutazioni

evidenziate nei capitoli precedenti, non si associano ad alcun arco. E' evidente che quando si considera un singolo punto, l'esigenza di ricondurlo all'arco più prossimo non permette di evidenziare il percorso nella sua complessità, dal suo inizio alla sua destinazione, e la pianificazione della strada da seguire. Il percorso ottimale scelto dall'individuo è definito anche, ad esempio, da comportamenti abitudinari, o dalla necessità di transitare in particolari luoghi, che rappresentano le personali àncore, da eventuali interruzioni stradali o cambi temporanei di sensi di marcia. Queste considerazioni hanno spinto ad analizzare i punti GPS non indipendenti l'uno dall'altro, ma nella loro globalità e generati da un' unica funzione. Anche i tratti che compongono una strada della mappa digitale sono considerati come unità costituenti lo stesso cammino che permette di raggiungere una destinazione, posto che partano da un punto di inizio. Se però il percorso rilevato è unico, le strade possibili possono essere molteplici: si tratta di confrontare, in modo globale, la rappresentazione delle possibili alternative già definite dalla mappa, con una curva particolare formata dai punti GPS rilevati.

Il problema quindi non è quello di condurre un punto o alcuni suoi 'vicini' su un tratto di strada, quanto quello di spostare globalmente il percorso su una delle diverse curve possibili della mappa digitale. Consideriamo ad esempio la figura 23 che riproduce la rete stradale, in termini di archi e nodi, relativi alla città di Cagliari: come si può notare non si tratta di una rete particolarmente complessa ma, ricondurre un percorso rilevato sulla rete stradale, non è certamente operazione di semplice risoluzione. Se si considera poi la figura 24, dove vengono rispettivamente considerati tutti i percorsi effettuati da un singolo soggetto in un arco di tempo (figura 24a), è evidente che l'operazione da compiere sarà quella di isolare uno ad uno i percorsi effettuati durante l'arco di tempo (punti in rosso nella Figura

24b), individuare il percorso della mappa che più si avvicina(Figura 234c) e procedere successivamente ad una verifica di convergenza dell'operazione effettuata (figura 24d).

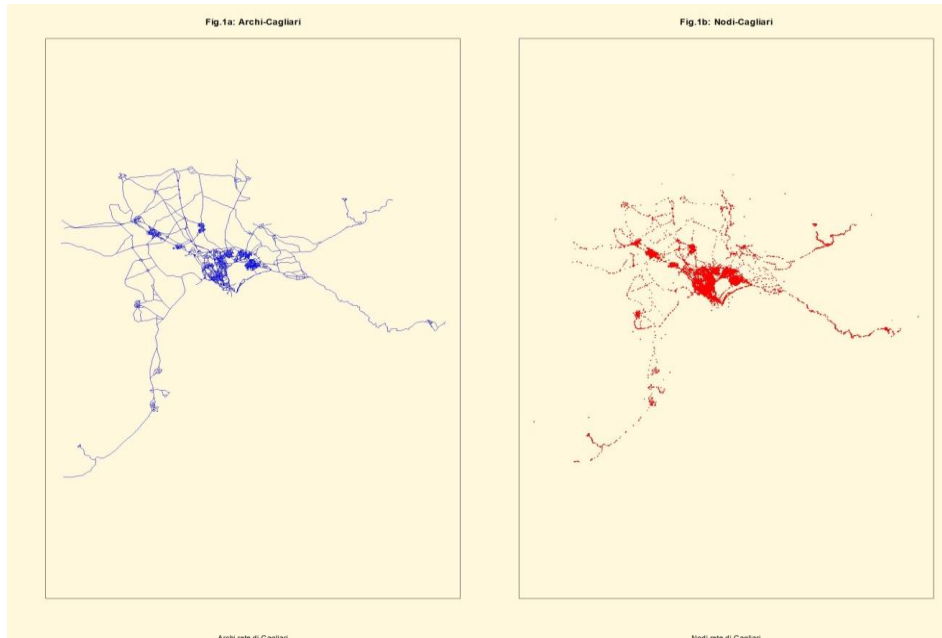


Figura 23: Archi (a) e nodi (b) della rete stradale di Cagliari

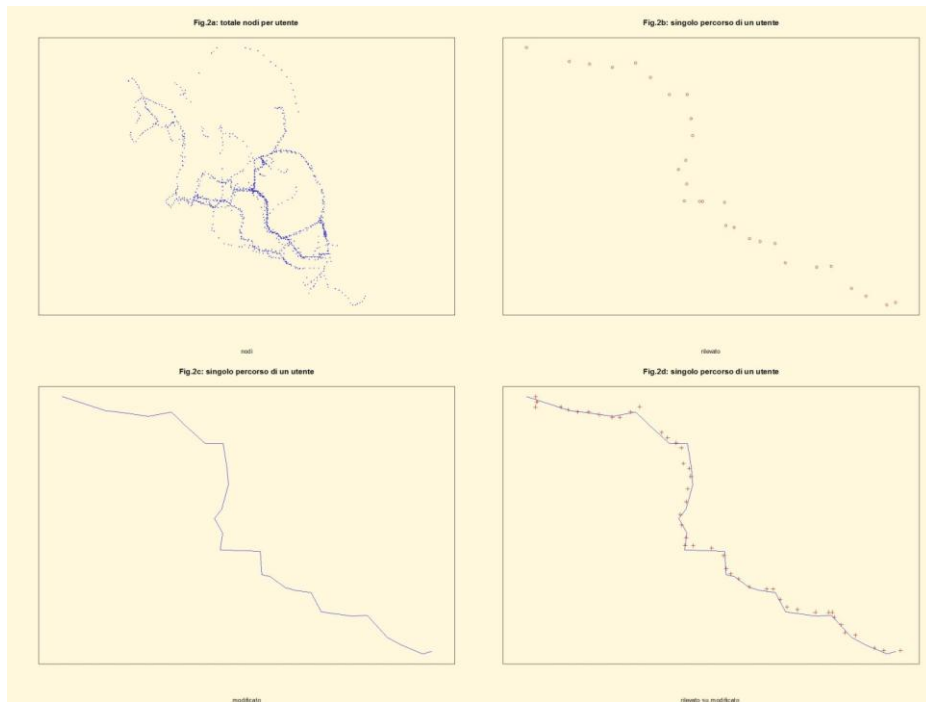


Figura 24: Percorsi di un individuo(a), punti GPS di un percorso (b), percorso (c), sovrapposizione (c)

E' evidente che non tutti i punti osservati si posizionano perfettamente sugli archi della strada individuata. Per comodità di esposizione si analizza il problema utilizzando un percorso formato da dati artificiali (figura 25a e b) e alcune strade della mappa digitale, anch'esse create per la circostanza. Nella figura 25c si rappresenta un potenziale percorso candidato e nella figura 25d si sovrappongono i due percorsi, in rosso il percorso GPS e in nero la strada della mappa. Se l'analisi visiva viene condotta per singolo punto, si nota che alcuni punti saranno più vicini alla strada, uno si sovrappone, ed altri si collocano lontano. Ma se l'analisi viene condotta con una visuale più ampia, ovvero globale, si nota che esiste uno scostamento complessivo che può essere considerato come una penalizzazione o un costo, che si sopporta nell'essere più distanti da una particolare strada.

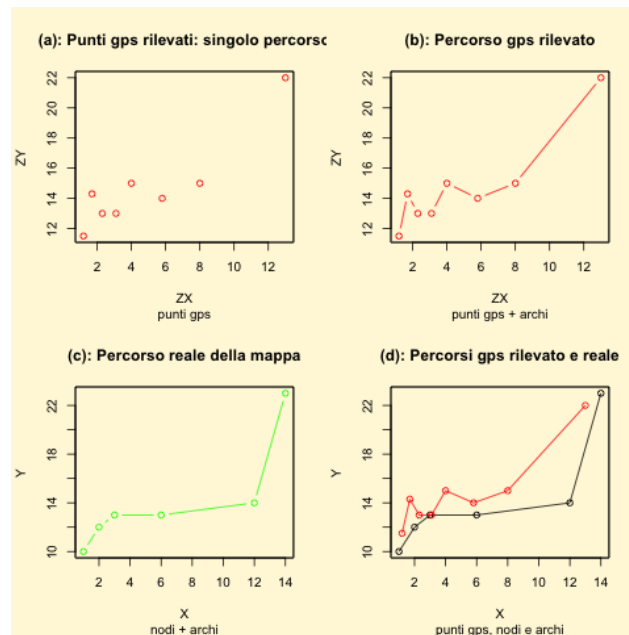


Figura 25: Singolo percorso GPS (a), spezzata del percorso (b), percorso reale (c), sovrapposizione (d)

La procedura che viene sviluppata si propone di individuare la quantità di tale penalizzazione e di definire, tra una rosa di strade candidate, a quale di queste è associato il minor costo. Se ad esempio non si conosce quale sia la strada da percorrere per giungere ad una particolare destinazione, un' informazione proveniente dall'esterno può far propendere per una strada piuttosto che per un'altra. Ma se durante il percorso si sbaglia direzione, si sarà costretti a ritornare sulla strada iniziale. Si consideri la strada come la spezzata in nero della figura 25d e quella in rosso il percorso per riprendere la strada corretta: ogni volta che ci si accorge dell'errore bisognerà variare direzione per ritornare sulla strada.

Ognuno di questi errori possono essere quindi visti come delle penalizzazioni o dei costi; quanto più un punto rilevato si discosta dal percorso vero tanto più la penalizzazione è alta; quanto più il punto è sulla strada vera, tanto più la penalizzazione è bassa o tende ad annullarsi.

Terminato il percorso, si saranno accumulate delle penalizzazioni pari alla somma di tutti i costi sopportati per ritornare sulla corretta strada. L'obiettivo sarà quello di accumulare la minore quantità di penalità.

4.2 Il windowing

Bisogna, però, individuare le possibili strade candidate; con l'aiuto delle coordinate spaziali acquisite tramite il ricevitore GPS è possibile individuare una finestra in cui ricadranno un certo numero di strade candidate. In figura 26 si riporta un percorso rilevato nell'indagine pilota e le strade che si ritrovano in tale area, mentre nella figura 27 si riporta una possibile zona intorno al percorso considerato, ovvero una finestra.

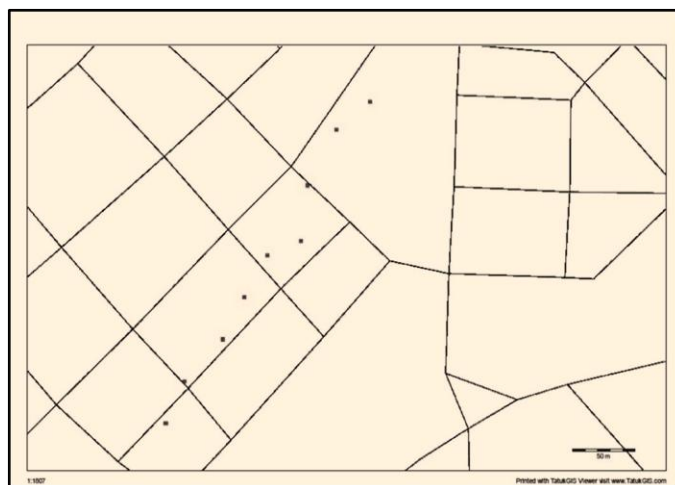


Figura 26: Un percorso rilevato: punti GPS

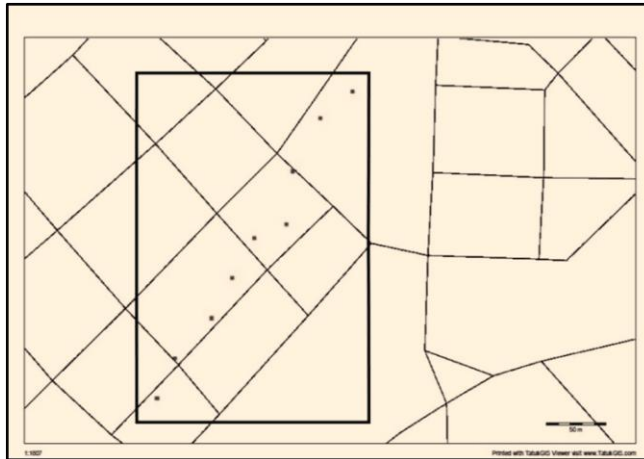


Figura 27: La finestra di analisi intorno al percorso GPS

Le strade che verranno prese in considerazione saranno unicamente quelle che ricadranno nella finestra; all'interno di questo insieme verranno successivamente considerate solo quelle che risultano coerenti con il percorso GPS considerato (coerenza rispetto alle informazioni topologiche).

Se il percorso rilevato dai punti GPS è particolarmente ampio o complesso da analizzare globalmente, si può procedere all'individuazione di finestre multiple, cioè finestre più piccole all'interno della finestra principale, dove applicare localmente la procedura (Figura 28).

L'individuazione della finestra di riferimento viene determinata dall'algoritmo in modo automatico, partendo proprio dalle coordinate dei punti GPS e considerando un valore di soglia che permette di aumentare o diminuire la grandezza della zona. Si consideri la Figura 29 in cui sia definito un percorso costituito dai punti A, B, C, D. Ogni punto originale GPS è spostato su una nuova posizione che ha coordinate tali da creare un poligono di interesse, contenente esattamente tutti i punti (chiamiamo questi nuovi punti AA, BB, CC, DD).

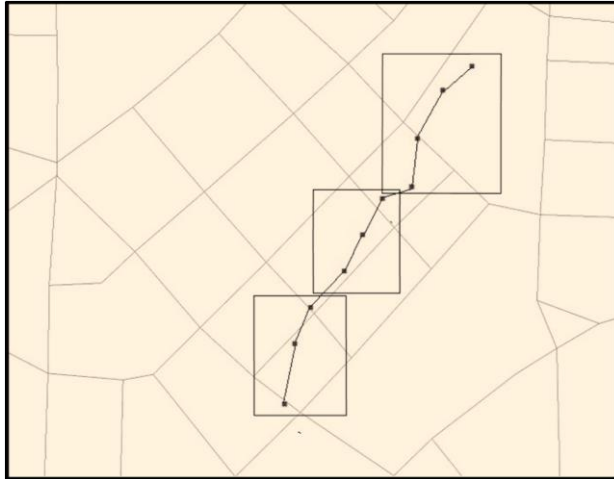


Figura 28: Il sub-windowing

Successivamente identifichiamo la finestra globale come una figura regolarizzata che contiene AA, BB, CC e DD e contemporaneamente tiene conto del parametro Δ (definito a priori), ottenendo i punti WA, WB, WC e WD che definiscono la finestra.

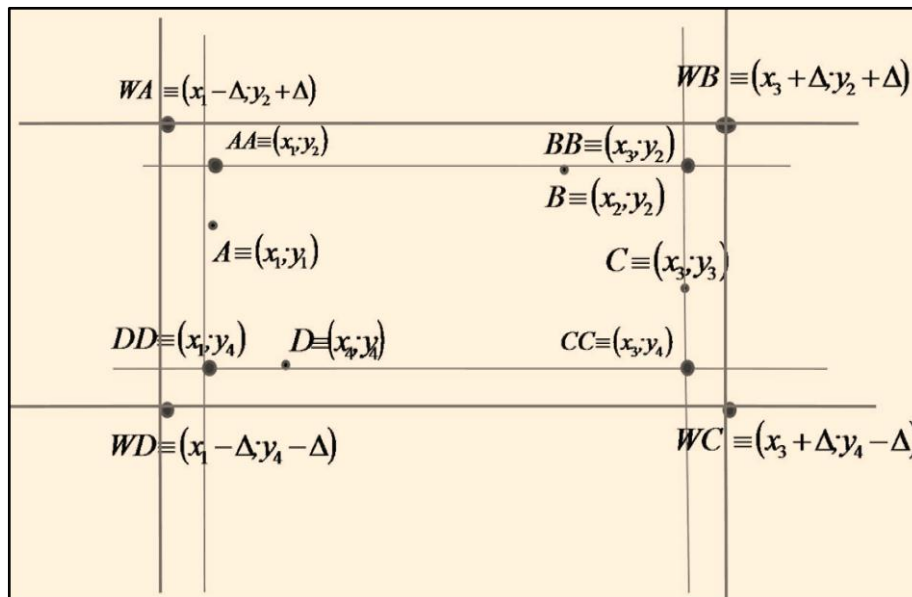


Figura 29: Creazione della finestra

Una volta effettuata questa operazione e definito il set dei percorsi candidati, ci si può, per comodità, svincolare dall'unità di misura della mappa e considerare un nuovo spazio di riferimento ove operare.

4.3 La definizione delle penalizzazioni

Bisogna calcolare le diverse penalizzazioni per poter operare una scelta del miglio candidato. Si consideri la figura 9a dove si riporta in rosso un punto GPS e in nero un tratto di strada. nella figura 9b viene evidenziato il percorso in più che si è compiuto dovendosi spostare dal nodo A al punto P; se si decide di ritornare sulla strada si hanno due alternative. La prima è ritornare indietro lungo lo stesso tratto, contando due volte il tragitto; oppure, ipotizzando che la direzione sia da sinistra verso destra, si compirà un percorso da P a B come mostra la figura 9d.; il percorso che permetterebbe di rientrare sul percorso correttamente (in questo caso alla lunghezza del segmento AP, andrebbe sommata la lunghezza del segmento PB). Ma questa non è l'unica scelta che si può compiere; infatti si potrebbe propendere, quando possibile, di riportarsi da P al segmento AB quando se ne presenti l'occasione. L'area del triangolo APB (come evidenziato in figura 9d) diventa quindi una possibile soluzione per ritornare sulla strada principale e può essere visto come il danno totale arrecato all'automobilista quando questi si è spostato dalla traiettoria AB al punto P. Analizzando sotto un aspetto diverso il problema, è come se il punto P appartenesse originariamente al segmento AB e poi "spostato" nella posizione P: l'area del triangolo APB è un costo che si è creato alla regolarità del segmento AB. Se ripetiamo questa operazione per gli n punti GPS che formano il nostro percorso e consideriamo non più un segmento ma un intero percorso formato da più segmenti, allora si avranno n triangoli che si possono costruire e n aree che possono essere calcolate.

La somma di queste n aree darà una misura della prossimità dell'intero percorso (figura 30).

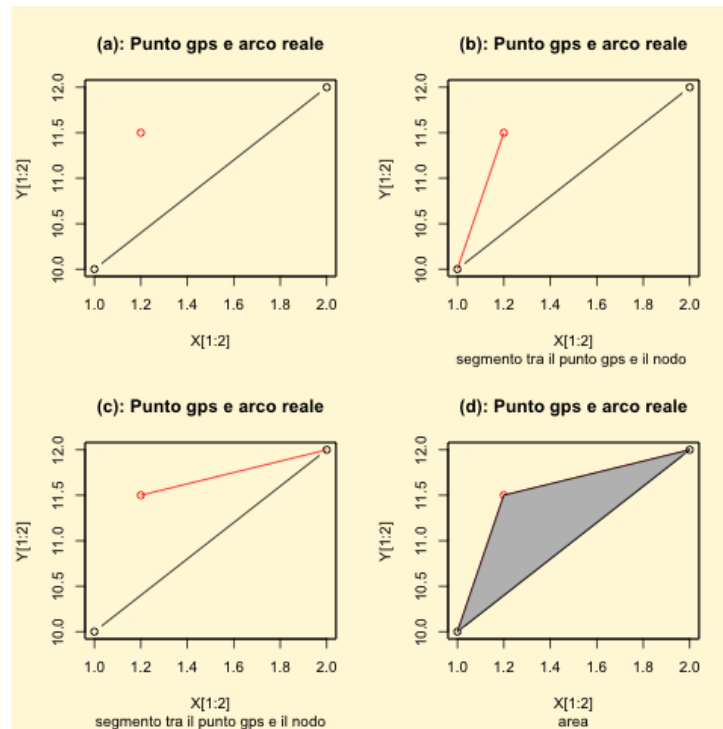


Figura 30:Costruzione dell'area

considerato ad una strada, appartenente all'insieme delle strade candidate. Può accadere che uno o più punti GPS appartengano ad un segmento e di conseguenza l'area calcolata sarà pari a zero, o molto più realisticamente prossima a zero. In questo caso la penalizzazione totale, data dalla somma di tutte le aree non incrementerà, ma rimarrà costante.

Ripetendo l'operazione per tutte le k strade nel set delle strade candidate, si avranno k valori relativi alla penalizzazione totale, derivante dal percorso rilevato e i k percorsi candidati. Individuata la curva che minimizza la penalizzazione totale, tutti i punti della curva formata dai punti GPS verranno ricondotti sugli archi (con proiezione ortogonale) appartenenti alla strada k^* scelta.

4.4 I diversi passi della procedura

Si può schematizzare il procedimento proposto come segue:

1. *Definizione dei parametri di funzionamento (soglia per la finestra di lavoro, numero massimo di percorsi candidati da considerare, numero massimo di sotto-finestre da considerare)*

In questa fase vengono inizializzati i parametri di funzionamento dell'algoritmo che possono variare in funzione della complessità della mappa, del numero di punti GPS e del numero massimo di sotto-finestre che si desiderano costruire in caso di percorsi molto lunghi o complessi.

2. *Acquisizione dei punti GPS rilevati e tracciamento di un percorso C sovrapposto ad una mappa digitale*

Questo è il primo passo da compiere ed è relativo alla selezione dei punti GPS che formano il percorso in esame. In questa fase il percorso viene proiettato su una mappa digitale al fine di poter pervenire ad una prima individuazione della zona di intervento.

3. *Acquisizione delle informazioni statiche e dinamiche presenti nel data base della mappa digitale*

Questa fase è particolarmente delicata in quanto permetterà di definire successivamente il set di possibili strade candidate.

4. *Definizione della finestra W globale sulla mappa*

In questa fase si definisce, seguendo l'approccio proposto e tenendo conto della soglia Δ , la finestra globale di riferimento, ovvero porzione di mappa attiva o di interesse.

5. *Eventuale individuazione di M sotto finestre ($w_i, i = 1, \dots, m, \dots, M$)*

Nel caso in cui il percorso sia molto lungo o complesso, può essere utile lavorare con finestre di dimensioni ridotte, al fine di ottimizzare il funzionamento dell'algoritmo, sia in termini di efficacia che di efficienza.

6. *Definizione del set O di percorsi reali, candidati come strade possibili ed appartenenti alla finestra W*

Definita la finestra principale e le eventuali sotto finestre, si procede, alla definizione di tutte le possibili strade che possono essere candidate. In questa fase l'algoritmo considera tutti i nodi della mappa digitale appartenenti alla finestra e costruisce tutti i possibili archi.

7. *Analisi della coerenza dei percorsi individuati appartenenti ad O e definizione dell'insieme Ω dei K percorsi reali candidati*

In questa fase, l'insieme O precedentemente considerato, viene depurato dalle incoerenze: tracciati gli archi, non tutti saranno coerenti con la curva C , in termini di coerenza topologica con il percorso C . Queste informazioni a corredo, unitamente ad informazioni esterne (sensi unici momentanei, interruzioni e deviazioni stradali temporanee, ecc) renderanno minore la cardinalità dell'insieme Ω e di conseguenza permetteranno di velocizzare la procedura.

8. *Per $1 \leq k \leq K$, costruzione dei triangoli tra i punti di C e la k -esima strada reale. Calcolo della misura totale di penalizzazione A_k*

In questa fase si costruiscono i triangoli, si calcolano le penalizzazioni parziali e quella globale. Il calcolo è basato sulla formula di Erone per il calcolo di aree, considerando nota le misure dei lati del triangolo. Si

tratta quindi a ben vedere di una operazione che porta ad una complessità computazionale lineare in n , rendendo l'algoritmo particolarmente performante.

9. *Individuazione del percorso reale k^* : $A_k = \min$*

Una volta iterata il punto 8 per tutte le K strade alternative, l'algoritmo seleziona quella con la penalizzazione globale minima.

10. *Riconduci i punti GPS di C sugli archi adiacenti del percorso reale k^**

L'ultima fase dell'algoritmo è il *map-matching*, ovvero ricondurre i punti GPS rilevati sugli archi della strada k^* individuata dalla procedura.

4.5 Un esempio con dati sperimentali

In questo paragrafo si mostra un esempio di funzionamento della procedura, focalizzando in particolare l'attenzione sulle fasi relative al calcolo della penalizzazione e alla selezione della strada scelta tra le possibili candidate. In figura 10a sono rappresentati i punti GPS rilevati e relativi ad un percorso effettuato. Si ipotizzi di avere definito l'insieme Ω del set delle strade candidate; in particolare si sono definite 4 strade appartenenti all'insieme: Nelle figure 31b-31e si definiscono le rappresentazioni grafiche del percorso e le quattro strade candidate, prese singolarmente; nella figura 31f si riassume le precedenti. Descritte le strade si procede nell'individuazione delle aree e si ricerca la penalizzazione per ognuna di loro. Nelle figure successive si rappresentano le aree formate congiungendo i punti GPS agli archi della prima utilizzando i nodi più vicini ai punti GPS.

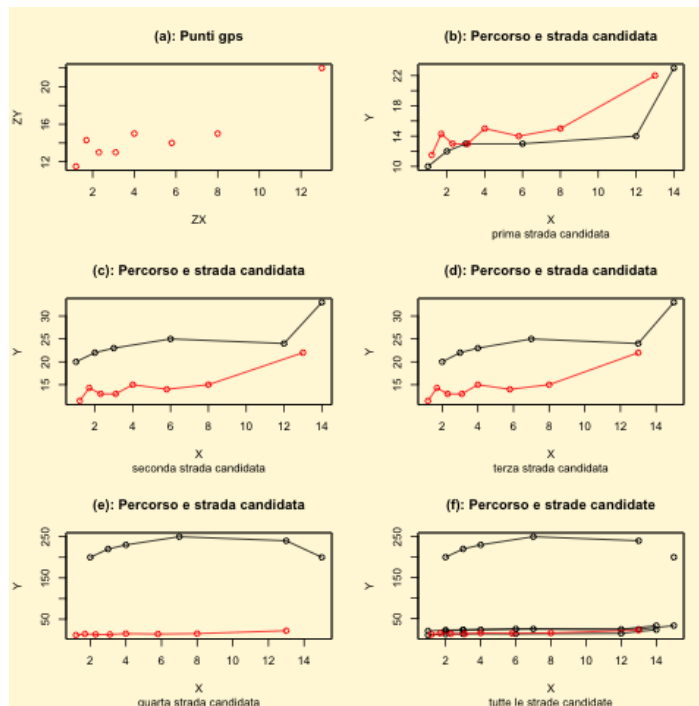


Figura 31: Percorso e strade candidate

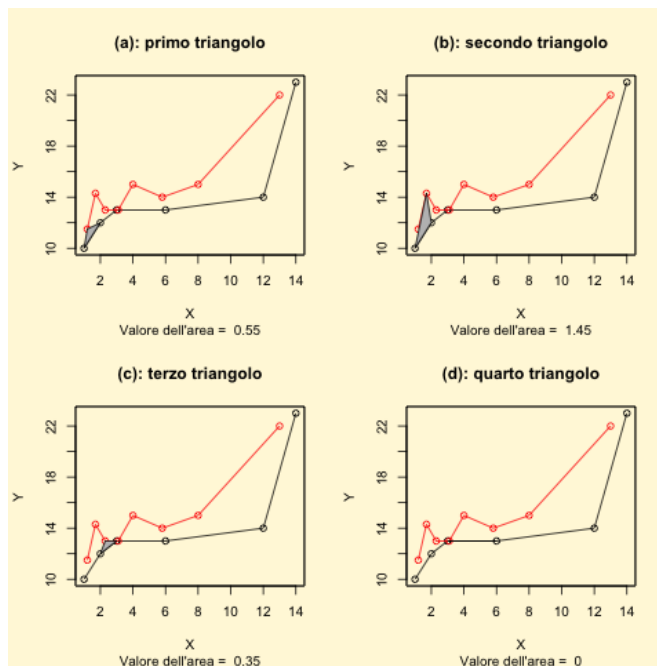


Figura 32: La costruzione dei triangoli per la prima strada

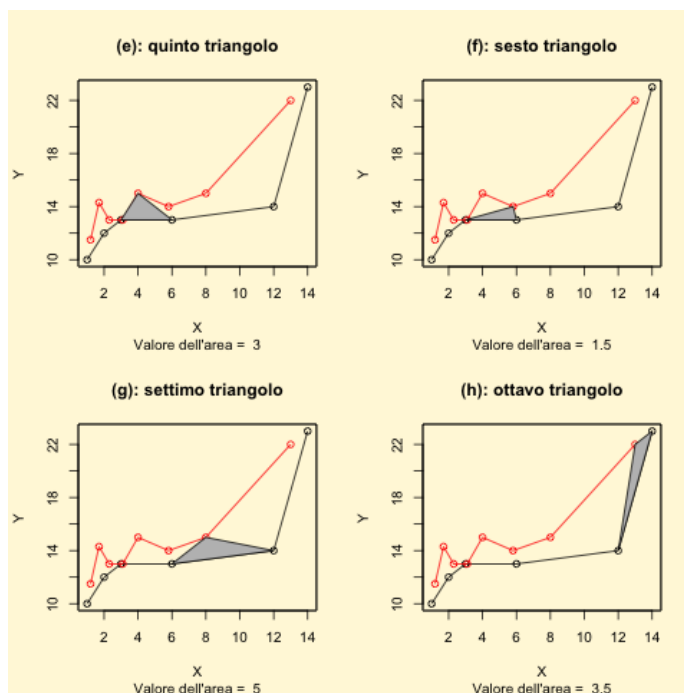


Figura 33: La costruzione dei triangoli per la prima strada

Al di sotto di ogni rappresentazione grafica è riportato il valore dell'area specificata. Come si vede dalla figura 32a il primo punto GPS del percorso non ricade sull'arco definito dai primi due nodi della strada presa in considerazione come candidata; viene costruito il triangolo che rappresenta una penalizzazione per il 'fuori percorso'. L'area del triangolo viene calcolata mediante la formula di Erone e il risultato parziale memorizzato; questa operazione viene iterata per tutti i punti del percorso (figure 32b – 33h). Alla fine del processori otterrà una misura globale data dalla somma di tutte le aree parziali. La stessa operazione viene effettuata per le rimanenti tre curve appartenenti al set Ω ; la rappresentazione grafica è rappresentata dalle figure successive.

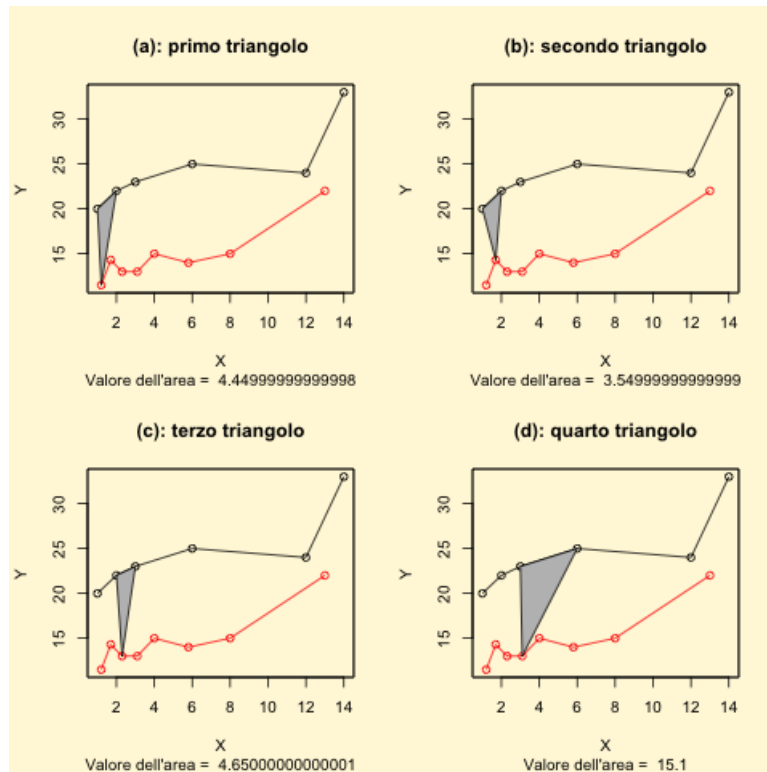


Figura 34: Iterazione per la costruzione delle aree per la seconda strada

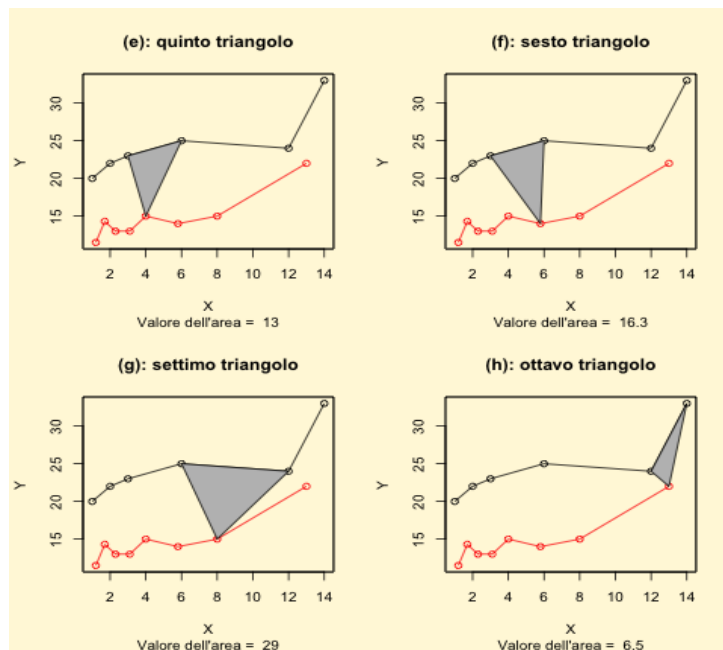


Figura 35: iterazione per la costruzione delle aree della seconda strada

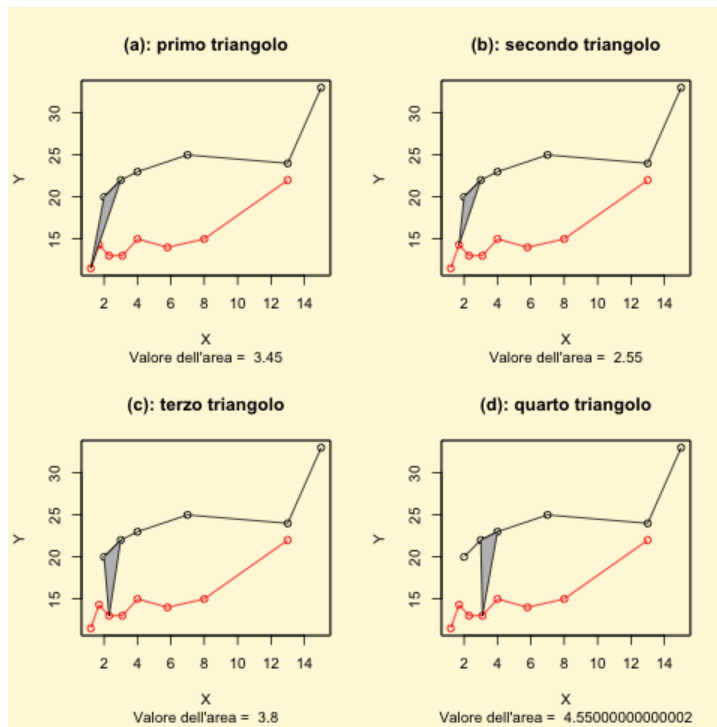


Figura 36: Iterazione per la costruzione delle aree per la terza strada

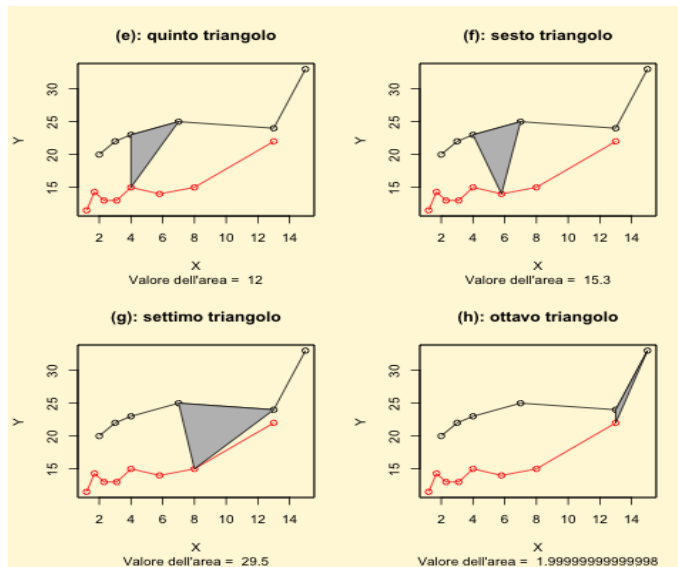


Figura 37: Iterazione per la costruzione delle aree per la terza strada

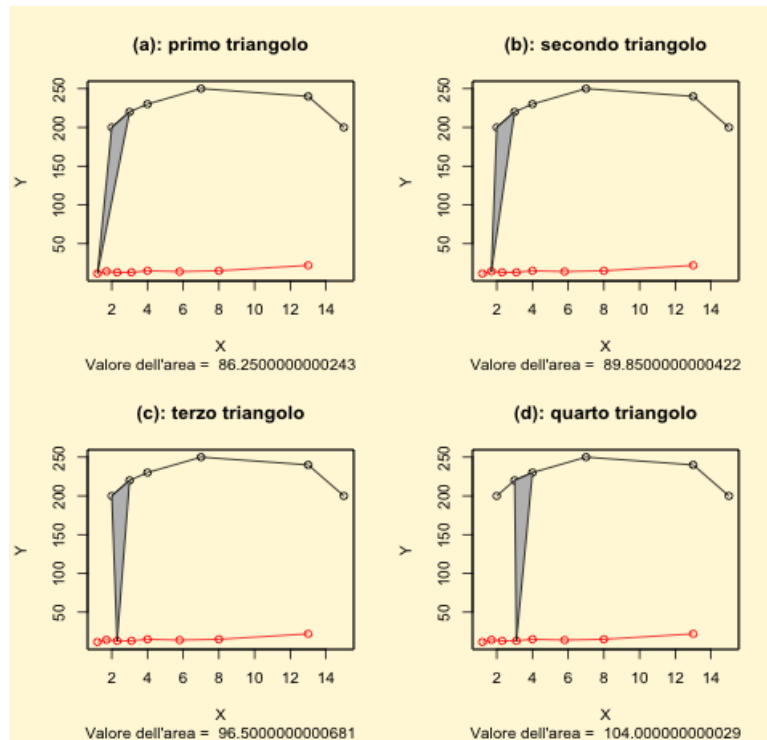


Figura 38: iterazione per la costruzione delle aree per la quarta strada

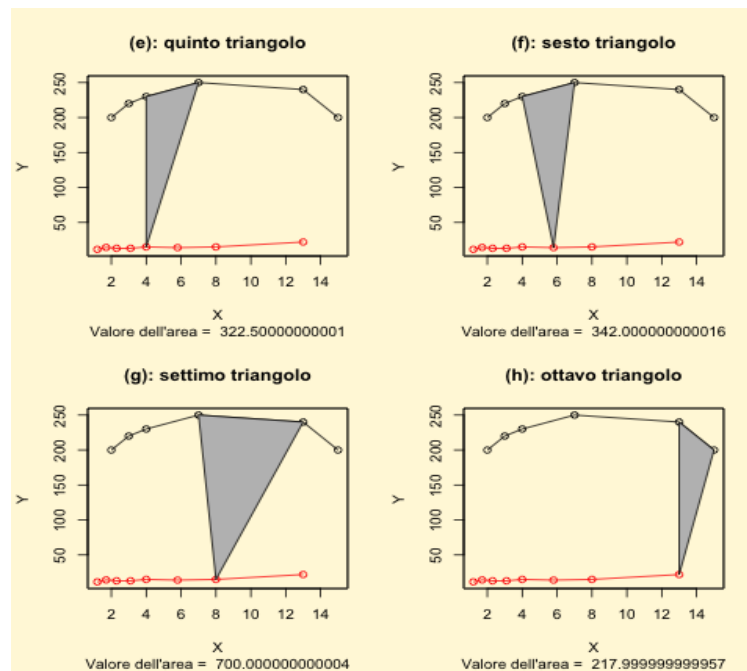


Figura 39: iterazione per la costruzione delle aree per la quarta strada

Alla fine del processo, l'algoritmo ha prodotto quattro misure di penalizzazione totale, che sono riassunte nella tabella 16:

Tabella 16: le penalizzazione delle 4 strade

<i>Percorso</i>	<i>Penalizzazione</i>
1	15.35
2	92.55
3	73.15
4	1959.10

Il percorso k^* che quindi minimizza la penalizzazione è il percorso numero 1.

A questo punto l'algoritmo ricostruisce i punti gps di C sugli archi della strada reale.

4.6 Considerazioni conclusive e possibili sviluppi di ricerca

Dopo aver illustrato il funzionamento della procedura proposta, si possono elencare ed analizzare alcuni aspetti della procedura stessa e dell'algoritmo di calcolo:

- 1) La metodologia proposta tende a coniugare da un lato l'approccio geometrico al *map matching* e dall'altro tiene conto della logica comportamentale. Infatti, considerando l'intero percorso si può definire un sistema che tenda a valutare anche il comportamento di un automobilista coerente.
- 2) L'algoritmo di calcolo risulta estremamente veloce in quanto è lineare ed è funzione unicamente delle coordinate geografiche dei punti GPS rilevati e le coordinate dei nodi della rete stradale.

- 3) L'algoritmo sfrutta le informazioni a corredo esistenti nella mappa digitale e può prendere in considerazione anche semplici mutamenti temporanei della rete stradale di riferimento (interruzioni, deviazioni momentanee, condizioni meteo, etc.)
- 4) Poiché la procedura genera un valore finale della funzione di penalizzazione per tutte le strade candidate e presenti nel set Ω , è possibile creare una graduatoria dei possibili candidati al *matching*. Questo significa che si può ottenere un ordinamento per similarità dei percorsi alternativi ed utilizzare in futuro questo approccio al problema come un simulatore di comportamenti individuali simili, rispetto a particolari circostanze ambientali (variazioni di sensi di marcia, chiusura temporanea di tratti stradali ,ecc).
- 5) Se si opera un'indagine nel tempo su un individuo, come nel caso della rilevazione mediante l'*activity locator*, la conoscenza delle abitudini del soggetto può definire un sistema di pesi probabilistici che viene in aiuto alla scelta della migliore strada candidata.
- 6) La procedura proposta è indipendente dalla mappa digitale di riferimento: quando si considera la finestra attiva per la definizione dell'insieme O e successivamente Ω , il percorso C e le strade candidate vengono considerate in un sistema di misura indipendente e quindi possono essere viste come misure normalizzate (eliminando gli effetti topologici).
- 7) Modifiche all'algoritmo di calcolo possono facilmente portare ad implementare strategie alternative per il calcolo della misura di penalizzazione. Infatti, possiamo considerare anziché triangoli, dei parallelogramma tra archi della mappa e archi formati da due punti GPS; è possibile considerare poligoni di penalizzazione formati

dalla congiunzione dei punti risiedenti sulla curva C e la strada candidata al matching.

A conclusione di questo capitolo si vuole sottolineare che la ricerca in questo settore ha ancora spazi per la proposizione di metodi sempre più efficienti ed efficaci per la risoluzione del problema. Molto può essere fatto per migliorare la proposta qui delineata, sia per quanto riguarda l'introduzione delle abitudini comportamentali degli automobilisti, sia per quanto riguarda l'introduzione di funzioni di penalizzazione differenti (come accennato nel punto 7 di questo paragrafo), che per le strategie di individuazione del *windowing* e *sub-windowing*.

Conclusioni

Durante lo svolgimento dell'indagine pilota svolta dal CRiMM dell'Università di Cagliari per analizzare i comportamenti di viaggio degli utenti della rete stradale e favorire modalità di spostamento alternative, si sono evidenziati alcuni problemi relativi al trattamento delle informazioni acquisite. La grande quantità di dati e informazioni spaziali e temporali ottenuta grazie all'istallazione di un'applicazione informatica su uno *smart- phone* dotato di ricevitore GPS, è stata utilizzata, tra l'altro, per poter ricostruire i reali percorsi osservati su un grafo della rete stradale della città e poter operare delle successive inferenze. E' in questo passaggio operativo che è sorta l'esigenza di analizzare e correggere la diversione dei punti posizione rilevati rispetto al tratto stradale rappresentato sulla mappa digitale (grafo di rete), ed eventualmente trovare una procedura automatica che limitasse la perdita di informazioni e dati e conducesse a ritrovare il percorso più vicino a quello vero.

L'articolazione del percorso di tesi si è svolto in prima fase con la metodica del *brain storming* che ha permesso di mettere in luce le molte interrelazioni esistenti tra i diversi campi disciplinari e suscitare curiosità ed interesse. Successivamente è stato necessario dare un ordine allo studio, all'approfondimento e all'analisi delle motivazioni alla base del crescente interesse dei ricercatori di voler ottenere informazioni corrette del tracciamento spaziale di percorsi stradali, e lo studio delle diverse procedure di *map matching* che nel tempo sono state proposte e sviluppate. Ciò ha permesso di realizzare un lavoro di tesi articolato nel campo della matematica e delle sue applicazioni in particolari settori

applicativi: pertanto si sono approfonditi i motivi della non sempre soddisfacente qualità dei dati raccolti con dispositivi GPS, l'importanza nell'ambito delle ITS dell'esatta proiezione e ricostruzione di un percorso reale su una mappa digitale, e come nel tempo si sono evolute le strategie di analisi. Sono stati individuati ed analizzati quegli algoritmi di *map matching* diventati fondamentali per lo sviluppo successivo di ulteriori strategie operative e il loro studio ha permesso di definire una proposta operativa a conclusione del percorso del ciclo di studio.

Il contributo metodologico tende a coniugare l'approccio geometrico alla risoluzione del problema e la conoscenza non solo delle caratteristiche dello spostamento ma anche le informazioni relative alla propensione di scelta delle alternative dell'individuo. L'individuazione delle possibili strade candidate al *map matching* viene realizzata mediante la creazione di uno spazio limitato, definito come finestra, in cui ricade una parte di rete stradale intorno al percorso rilevato del veicolo. Successivamente, per operare una scelta tra il set di strade candidate si considera la penalizzazione (l'area dei triangoli costruiti) del percorso rispetto ad ogni singola strada candidata e si opta per quella strada che permette di raggiungere la minore penalità totale. L'algoritmo sviluppato in R oltre ad individuare la finestra di esame, permette di calcolare le aree dei triangoli che si creano tra il percorso ricostruito attraverso i punti posizione osservati e le strade rappresentate dal grafo di rete. Definiti i lati come quei segmenti che si formano tra i punti GPS e i nodi degli archi di strada prossimi al punto, si utilizzano le coordinate geografiche e si sommano consecutivamente tutte le aree dei triangoli formatisi tra tutto il percorso e ciascuna intera strada candidata.

I risultati sperimentali ottenuti dall'applicazione di questa strategia sembrano incoraggianti anche se il contributo metodologico sviluppato in

questo lavoro può certamente essere ampliato e sviluppato con l'introduzione di vincoli sia topologici che comportamentali che definiscano la funzione di scelta della penalizzazione e con la definizione del fattore di soglia per la definizione delle finestre di indagine, rispetto a dei parametri caratterizzanti sia la complessità della rete stradale che quella del territorio.

Appendice A

Il filtro Kalman

Il Filtro di Kalman viene ideato alla fine degli anni '50 nell'ambito del programma spaziale americano. La motivazione applicativa che portò allo sviluppo del filtro fu la necessità di creare algoritmi che potessero determinare in maniera veloce e ricorsivamente la posizione e la velocità dei satelliti e delle navicelle spaziali, ovvero predire in tempo reale la traiettoria di un corpo in movimento, aggiornando la stima all'istante precedente in base alle nuove osservazioni e di un modello dinamico a priori. Teoricamente è uno strumento per stimare lo stato di un sistema lineare perturbato da rumore, basandosi su osservazioni linearmente dipendenti dallo stato e corrotte da rumore. Il filtro Kalman non è un filtro tradizionale ma un programma ricorsivo che si basa su informazioni in tempi differenti che elimina il rumore delle misure per costruire le stime.

Sia dato il seguente esempio: si consideri un veicolo in movimento le cui grandezze, posizione p e velocità v , si definiscono lungo la direzione x che y :

$$\begin{bmatrix} p_x \\ p_y \\ v_x \\ v_y \end{bmatrix}$$

Si vuole effettuare una stima della posizione del veicolo. In Figura A. 1 è riportata la rappresentazione della stima della posizione e della velocità

del veicolo, ipotizzando che: $\hat{x}_{k/k}$ è la stima dello stato all'istante k date le misure fino a k ; $\hat{x}_{k+1/k}$ è la predizione dello stato all'istante $k+1$ date le misure fino a k ; $\hat{x}_{k+1/k+1}$ è la stima ottima dello stato all'istante $k+1$ date le misure fino a $k+1$; y_{k+1} è la misure fornita dal ricevitore all'istante $k+1$. Le ellissi rappresentano l'incertezza, ovvero sono regioni di confidenza della stima; la stima $\hat{x}_{k+1/k+1}$ si troverà certamente sul segmento che congiunge $\hat{x}_{k+1/k}$ e y_{k+1} nell'ellisse centrata in $\hat{x}_{k+1/k+1}$ e dimensione proporzionale a $P_{k+1/k+1}$ ovvero alla posizione al tempo $k+1$ date le misure fino a $k+1$.

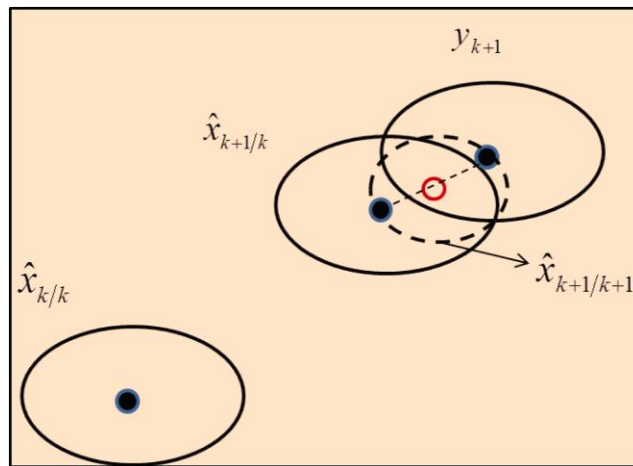


Figura A. 1: Regioni di confidenza della stima

Sia dato il seguente sistema lineare casuale a tempo discreto:

$$\begin{aligned}
 x_{k+1} &= Ax_k + w_k & \text{con } x_k \in \mathfrak{R}^n, y_k \in \mathfrak{R}^m, A \in \mathfrak{R}^{n \times n}, C \in \mathfrak{R}^{m \times n} \\
 y_k &= Cx_k + v_k & w_k = N(0, Q), v_k = N(0, R), x_0 = N(\bar{x}_0, P_0) \text{ ed } i.i.d. \\
 & & \text{con } Q, R, P_0 \geq 0 \text{ e } k \text{ istanti temporali}
 \end{aligned}$$

Definiamo le seguenti quantità:

$$\hat{x}_{k+1/k+1} = E[x_{k+1}|y_0, \dots, y_{k+1}] = E[x_{k+1}|Y^k] \text{ dove}$$

$$Y^k = (y_k, \dots, y_1, y_0) \in \mathfrak{R}^{(k+1)m}$$

$$P_{k+1|k+1} = E\left[(x_{k+1} - \hat{x}_{k+1|k+1})(x_{k+1} - \hat{x}_{k+1|k+1})^T | y_0, \dots, y_{k+1}\right]$$

con $P_{k+1|k+1} \geq 0$, varianza dell'errore della stima. Allora $\hat{x}_{k+1/k+1}$ viene definito come filtro Kalman. Se $k=h+N$ allora $\hat{x}_{k+1/k+1}$ è definito come predittore a N passi, mentre se $k=h-N$ allora $\hat{x}_{k+1/k+1}$ è visto come interpolatore (*smoother*) a N passi.

Si vuole trovare un'espressione ricorsiva che renda facile e veloce l'applicazione del filtro ad ogni passo.

L'espressione generica $E[X|Y]$ è definita come stimatore a minima varianza o stimatore ottimo in media quadratica. Esso è facilmente calcolabile se X e Y sono due variabili casuali congiuntamente gaussiane e

sono note μ_X, μ_Y , e $\begin{bmatrix} \sum XX & \sum XY \\ \sum YX & \sum YY \end{bmatrix}$.

Consideriamo $X = x_{k+1}|Y^k$ e $Y = y_{k+1}|Y^k$ per cui si avrà:

$$\begin{aligned} \mu_X &= E[x_{k+1}|Y^k] \\ \mu_Y &= E[y_{k+1}|Y^k] \\ P_{k+1|k} &= \sum_{XX} = \text{Var}[x_{k+1}|Y^k] \\ \sum_{YY} &= \text{Var}[y_{k+1}|Y^k] \\ \sum_{XY} &= \sum_{YX}^T = \text{Cov}[x_{k+1}, y_{k+1}|Y^k] \end{aligned} \quad [1]$$

Se si calcolano le forme esplicite si avranno le espressioni dello stimatore ottimo $\hat{x}_{k+1/k+1}$ e la varianza dell'errore della stima $P_{k+1/k+1}$

$$E[X|Y] = \hat{x}_{k+1/k+1} = \hat{x}_{k+1/k} + \sum_{XY} \sum_{YY}^{-1} (y_{k+1} - \hat{y}_{k+1|k})$$

$$P_{k+1/k+1} = \sum_{X|Y} = \sum_{XX} - \sum_{XY} \sum_{YY}^{-1} \sum_{YX}$$

Se si calcolano in forma esplicita le [1], dopo alcuni passaggi si avranno:

$$\begin{aligned} \mu_X &= E[x_{k+1}|Y^k] = \hat{x}_{k+1/k} = A\hat{x}_{k/k} \\ \mu_Y &= E[y_{k+1}|Y^k] = C\hat{x}_{k+1/k} = C\hat{x}_{k/k} \\ \sum_{XX} &= P_{k+1|k} = \text{Var}[x_{k+1}|Y^k] = AP_{k|k}A^T + Q \\ \sum_{YY} &= \text{Var}[y_{k+1}|Y^k] = CP_{k+1|k}C^T + R \\ \sum_{XY} &= \sum_{YX}^T = \text{Cov}[x_{k+1}, y_{k+1}|Y^k] = P_{k+1|k}C^T \end{aligned}$$

Sostituendo le forme esplicite nell'espressione dello stimatore ottimo e nella varianza dell'errore di stima si otterranno le equazione ricorsive del filtro Kalman:

$$E[X|Y] = \hat{x}_{k+1/k+1} = \hat{x}_{k+1/k} + K_{k+1|k+1} (y_{k+1} - C\hat{x}_{k+1/k})$$

$$P_{k+1/k+1} = P_{k+1|k} - P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}CP_{k+1|k}$$

con

$$K_{k+1|k+1} = P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}$$

con condizioni iniziali :

$$\hat{x}_{0|-1} = E[x_0] \quad P_{0|-1} = \text{Var}[x_0]$$

Se si ipotizza che il sistema sia dinamico a tempo discreto ma con x_{k+1} e y_k non lineari e con rumori di processo non necessariamente variabili casuali gaussiane, allora il filtro Kalman classico non può essere applicato. L'algoritmo prende allora il nome di Filtro Kalman Esteso (EKF) e si basa sull'applicazione del filtro Kalman al sistema linearizzato istante per istante. Si ipotizza che la mappa z sia $z = g(\zeta)$ con ζ una variabile casuale a distribuzione $p_\zeta(\cdot)$ nota. Inoltre, $\zeta \in \mathfrak{R}^m, z \in \mathfrak{R}^n$ e $g: \mathfrak{R}^m \rightarrow \mathfrak{R}^n$ una mappa non lineare nota. Il procedimento del EKF approssima p_ζ ad una variabile gaussiana con media $\hat{\zeta}$ e varianza P_ζ e successivamente approssima la mappa g con la sua linearizzazione. In tal modo la variabile z può essere approssimata da una variabile gaussiana con media \bar{z} e varianza P_z .

Il EKF si basa, quindi, su ipotesi di gaussianità e piccola varianza dei rumori ed in genere non permette di avere garanzia di stabilità; è possibile che per lunghi tratti il filtro esteso funzioni correttamente ma potrebbe divergere irrimediabilmente.

Appendice B

La ricerca bidirezionale del percorso più breve

Dato un sistema di interconnessione di nove nodi numerati da 1 a 9, tra di loro collegati, si vuole trovare la strada più veloce dal nodo origine a quello di destinazione. Per fare ciò si esaminano tutte le strade che partono dal punto 1 e arrivano al punto 9. Si acquisiscono due valori per ogni nodo, il suo numero x e la distanza dal punto 1 o dal punto 9. Successivamente ogni nodo verrà espanso con i punti adiacenti connessi; questa espansione avviene alternativamente partendo da nodo 1 e dal nodo 9 (Figura A. 2).

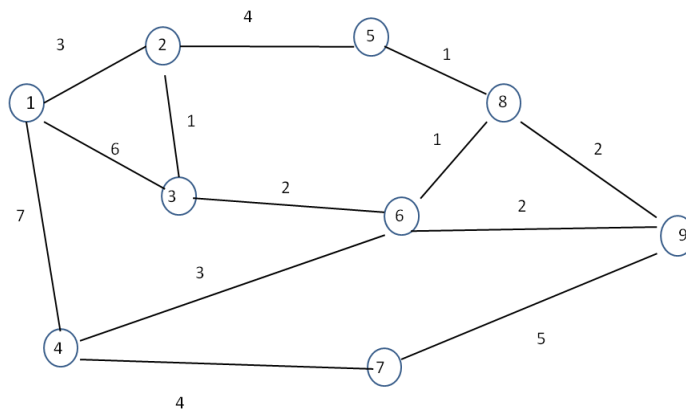


Figura A. 2: Possibili strade e distanze tra i nodi 1 e 9

Nella Figura A. 3 vengono ripercorsi tutti i passaggi finali dell'espansione.

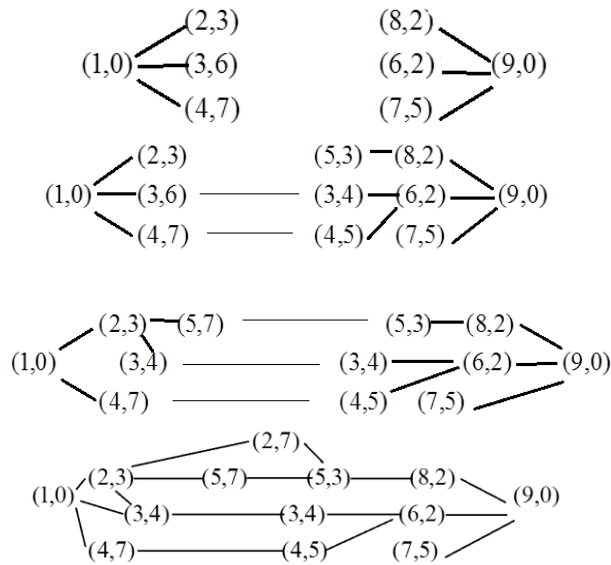


Figura A. 3: Espansione del percorso dal nodo 1 al nodo 9

Si può notare che nel secondo schema per arrivare al nodo 6 si è considerato unicamente il percorso 9-6, con lunghezza 2, e non 9-8-6, con lunghezza 3, in quanto la lunghezza del secondo percorso è superiore. Così come nel terzo schema per arrivare dal nodo 1 al nodo 3 si perde il passaggio 1-3, con lunghezza 6, ma si crea il percorso 1-2-3, con lunghezza 4. L'algoritmo consiste in due passaggi ciclici che si arrestano quando la strada ottimale è stata generata: il primo seleziona le strade da espandere ed espande le strade selezionate, il secondo si accerta che sia stata trovata la strada più veloce.

Sono da sottolineare due effetti di tale procedura che non rendono il metodo così efficiente: il primo riguarda lo sforzo computazionale che si rivolge, nelle singole iterazioni, alla ricerca del percorso minimo in cui sono presenti le lunghezze minime degli archi. Nell'iterazione viene analizzato il percorso fino al suo nodo minimo

e quindi la visione è concentrata su quella particolare area fino a esaminarla approfonditamente. Il secondo effetto è il criterio secondo cui le ricerche che partono dai due nodi estremi si arrestano. Utilizzando il criterio proposto da Nicholson il risultato che si ottiene è inferiore a quello che si otterrebbe con un processo unidirezionale, in quanto solo il ramo proveniente da uno dei due nodi estremi potrebbe essere completamente espanso, senza che l'incontro avvenga in un punto intermedio.

Il metodo del sotto- obiettivo prevede l'individuazione di un nodo intermedio o di un arco collocato tra il nodo di origine e quello di destinazione, grazie al quale poter identificare il percorso più breve. E' possibile anche in questa procedura, prevedere la decomposizione in due problemi: uno per trovare il percorso più veloce tra il nodo origine e il sotto obiettivo, l'altro tra questo e il nodo destinazione. L'efficienza del metodo dipende dal numero dei sotto obiettivi e dalla loro posizione: maggiore è il loro numero e la loro collocazione uniforme tra i due nodi origine-destinazione maggiore sarà l'efficienza computazionale. Però la soluzione trovata non è necessariamente quella ottimale, potendo dipendere dalle caratteristiche topologiche dell'infrastruttura stradale.

Appendice C

La logica fuzzy

Se il modello booleano definisce un criterio binario per la valutazione di una proposizione, la logica fuzzy si basa su un modello che non contempla unicamente proposizioni che assumono i valori di verità *vero* o *falso*, ma estende il modello al caso di forme parziali di ordinamento: si estende il modello insiemistico ad una logica che prevede un terzo valore, definito come *indeterminato*. Si introduce per questo la nozione di grado di appartenenza, associato a ciascun elemento di un insieme, che varierà tra i valori 0 e 1 . In una logica a tre valori i gradi di appartenenza sono definiti con $1, 0$ e $\frac{1}{2}$ in modo da descrivere delle appartenenze sfumate o *fuzzy*.

Oltre alla logica a tre valori son state definite anche logiche ad n valori di verità. Per ogni n dato i valori associati sono di solito numeri razionali appartenenti all'intervallo $[0,1]$. L'insieme di valori di verità possono essere interpretati come gradi di verità, così definiti:

$$T_n = \left\{ 0, \frac{1}{n-1}, \frac{2}{n-1}, \dots, \frac{n-1}{n-1} \right\}$$

Lukasiewicz è stato il primo a proporre le seguenti primitive per la logica ad n valori (L_n):

$$\begin{aligned} \bar{a} &= 1 - a \\ a \wedge b &= \min(a, b) \\ a \vee b &= \max(a, b) \\ a \Rightarrow b &= \min(1, 1 + b - a) \\ a \Leftrightarrow b &= 1 - |a - b| \end{aligned}$$

Le diverse logiche a più valori hanno la loro controparte nella teoria dei *fuzzy set* che forma il nocciolo della logica fuzzy. Essa viene sviluppata utilizzando gli operatori di complemento, unione e intersezione. Un *fuzzy set* può essere visto come un'estensione di un set classico: se X è l'universo in considerazione e x i suoi elementi allora un *fuzzy set* A in X è definito come:

$$A = \{x, \mu_A(x) \mid x \in X\}$$

La funzione caratteristica discriminazione (*Membership Function*) è $\mu_A(x): X \rightarrow \{0,1\}$ che associa ad ogni elemento di $x \in X$ un numero reale $\mu_A(x)$ nell'intervallo $[0,1]$. Date due funzioni caratteristiche A e B allora:

$$\begin{aligned} \mu_{\bar{A}}(x) &= 1 - \mu_A(x) \\ \mu_{A \cup B}(x) &= \max(\mu_A(x), \mu_B(x)) \\ \mu_{A \cap B}(x) &= \min(\mu_A(x), \mu_B(x)) \end{aligned}$$

$\mu_{\bar{A}}(x)$ è specificata dalla funzione $c: [0,1] \rightarrow [0,1]$ che assegna un valore $c(\mu_A(x))$ ad ogni grado di appartenenza $\mu_A(x)$. Questo valore può intendersi come grado di appartenenza dell'elemento x nel *fuzzy set*, che rappresenta la negazione di A . Si definiscono le seguenti condizioni:

$c(0) = 1$ e $c(1) = 0$	condizione di contorno
$c(a) \geq c(b) \quad \forall a, b \in [0,1] \quad a < b$	monotonia non crescente
c è una funzione continua	continuità
$c(c(a)) = a \quad \forall a \in [0,1]$	involutione

Nella Figura A. 4 si evidenzia l'insieme A con la linea rossa e la sua negazione con la linea nera tratteggiata.

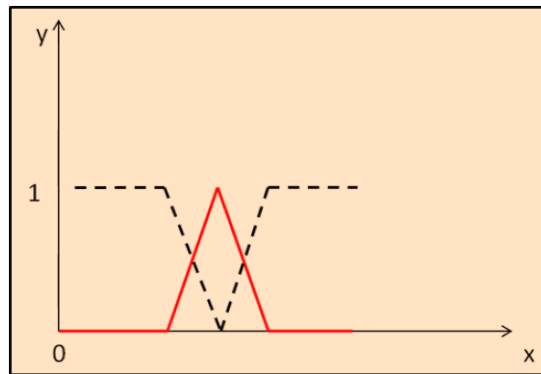


Figura A. 4: L'insieme A e la sua negazione

$\mu_{A \cup B}(x)$ è specificata dalla funzione $u: [0,1] \times [0,1] \rightarrow [0,1]$ per ogni elemento x dell'insieme universo e restituisce il grado di appartenenza dell'elemento appartenente all'insieme $A \cup B$. Si definiscono le seguenti proprietà:

$u(0,0) = 0$	$u(0,1) = u(1,0) = u(1,1) = 1$	condizioni di contorno
$u(a,b) \leq u(a',b')$	$\forall a \leq a' \text{ e } b \leq b'$	monotonia
u è continua		continuità
$u(a,a) = a$		idempotenza
$u(u(a,b),c) = u(a,u(b,c))$		associatività
$u(a,b) = u(b,a)$		commutatività

La Figura A. 5 definisce l'unione degli insiemi fuzzy A e B: la linea rossa rappresenta l'insieme fuzzy A, quella azzurra l'insieme fuzzy B e le linee non tratteggiate definiscono l'insieme unione è la seguente:

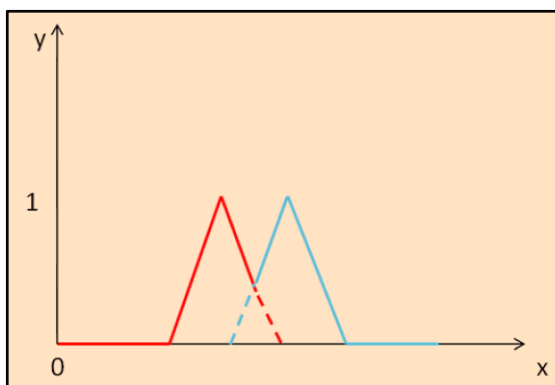


Figura A. 5: L'unione degli insiemi fuzzy A e B

$\mu_{A \cap B}(x)$ è specificata dalla funzione $i : [0,1] \times [0,1] \rightarrow [0,1]$ che restituisce il grado di appartenenza dell'elemento appartenente all'insieme $A \cap B$. Si definiscono le seguenti proprietà:

$i(1,1) = 1$	$i(0,1) = i(1,0) = i(0,0) = 0$	condizioni di contorno
$i(a,b) \leq i(a',b')$	$\forall a \leq a' \text{ e } b \leq b'$	monotonia
i è continua		continuità
$i(a,a) = a$		idempotenza
$i(i(a,b),c) = i(a,i(b,c))$		associatività
$i(a,b) = i(b,a)$		commutatività

Analogamente al grafico successivo, si può rappresentare l'insieme intersezione di due insiemi fuzzy, definendo la linea nera come la parte rappresentante l'insieme intersezione come nella Fig A. 6:

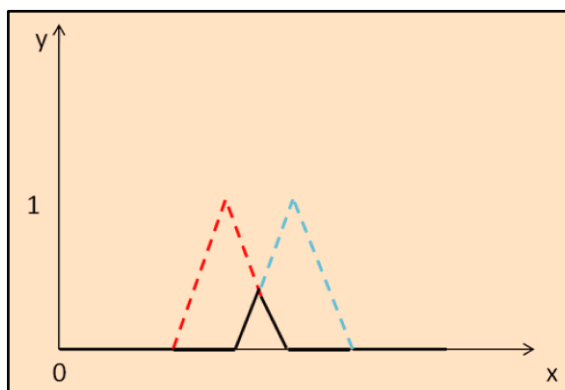


Fig A. 6: L'intersezione tra gli insiemi fuzzy A e B

Nella logica fuzzy il principio di contraddizione e del mezzo escluso non valgono, ovvero:

$$A \cup \bar{A} \neq X$$

$$A \cap \bar{A} \neq \emptyset$$

Queste circostanze si possono rappresentare graficamente considerando il risultato delle due operazioni con la linea nera continua (Figura A. 7):

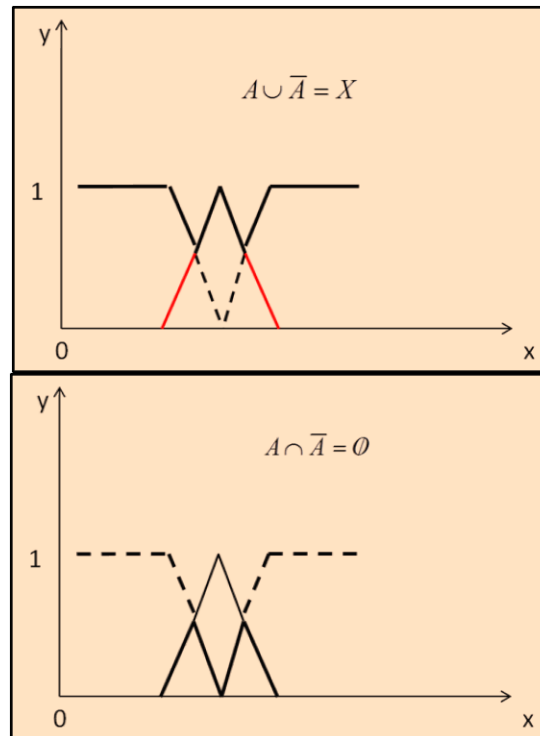


Figura A. 7: Principio di contraddizione e del mezzo escluso dell'insieme fuzzy A e della sua negazione

Sui fuzzy set si può definire una funzione che genera operazioni di aggregazione, per poter combinare vari insiemi e produrne solo uno. La funzione applicata a n fuzzy set A_1, A_2, \dots, A_n definiti su X produce un ulteriore set aggregato A , ovvero:

$$\mu_A(x) = h(\mu_{A_1}(x), \mu_{A_2}(x), \dots, \mu_{A_n}(x)) \quad \forall x \in X.$$

Tale funzione di aggregazione h è definita come $h: [0,1]^n \rightarrow [0,1]$ e deve rispettare almeno i

seguenti assiomi:

$h(0,0,\dots,0) = 0$	$h(1,1,\dots,1) = 1$	condizioni di contorno
$h(a_i i \in N_n) \geq h(b_i i \in N_n)$	se	
$a_i \geq b_i \quad \forall (a_i i \in N_n), (b_i i \in N_n)$	$a_i, b_i \in [0,1]$	monotonia non decrescente
h è una funzione continua		continuità
$h(a_i i \in N_n) = h(a_{p(i)} i \in N_n)$	\forall permutazione p su N_n	simmetria

Quando esiste un solo valore di x per cui la funzione di appartenenza è positiva, allora l'insieme fuzzy è detto *singleton*.

Alcune tra le più semplici funzioni *FM* sono la funzione triangolare, quella trapezoidale e la curva di Gauss, riportate di seguito (Figura A. 8).

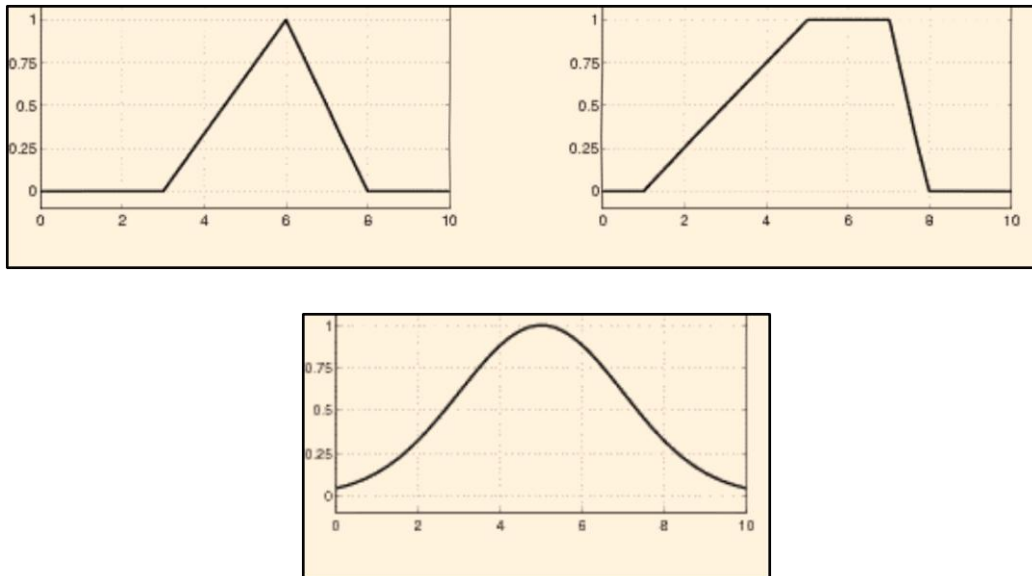


Figura A. 8: Alcune Membership Function

Il modello Fuzzy

Un modello fuzzy (*Model Fuzzy*) consiste in un insieme di regole deduttive che legano le determinanti in entrata e in uscita. Queste si basano su predicati e connessioni di tipo *IF-THEN*; la prima parte della regola è detta antecedente e la seconda conseguente. I modelli linguistici si basano sulle regole classiche di inferenza fuzzy; il modello più utilizzato è quello che prende il nome da E. Mamdani che utilizzò il metodo di inferenza fuzzy per la prima volta nei sistemi di controllo. Il modello prevede un antecedente e conseguente che vengono descritti mediante *fuzzy sets* e rappresentano concetti linguistici.

I modelli di Sugeno definiscono gli antecedenti che implicano dei conseguenti funzionali, ovvero deterministici:
IF{*Antecedent e*}*THEN*{*funzionale*}.

In particolare si può definire la successione di regole per un sistema di inferenza fuzzy, in quanto più utilizzato e supportato da software applicativi, come il MATLAB, che ne rendono semplice la valutazione. Il primo passo per creare un processo di inferenza fuzzy è la cosiddetta fuzzificazione degli inputs degli antecedenti, ovvero la trasformazione mediante la funzione caratteristica o MF in grado di appartenenza. Il secondo passo è l'applicazione dell'operatore fuzzy agli inputs fuzzificati per ottenere un singolo valore che sarà il risultato dell'operatore fuzzy.

Dopo aver assegnato ad ogni regola un proprio peso viene applicato l'operatore di implicazione, che produce un fuzzy set rappresentato da una funzione caratteristica che pesa appropriatamente le caratteristiche linguistiche. Il conseguente è riorganizzato usando una funzione associata all'antecedente; l'input di questa fase è un singolo

numero dato dall'antecedente, mentre l'output è un fuzzy set. I metodi utilizzati per implementare questa regole sono quelle valide per la *MF*, ovvero che $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$ e $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$. Se le regole sono più di una si formeranno tanti *fuzzy sets* quante sono le condizioni; per questo il quarto passo consiste nell'aggregazione dei *sets* per ottenerne uno unico; anche in questo caso valgono le uguaglianze precedenti. Infine l'ultimo passo è il processo di *defuzzificazione* ovvero la generazione di un singolo numero (*crisp*). Anche in questo circostanza è possibile utilizzare diversi metodi come il metodo del centroide, la media del massimo valore del set di output, il massimo più elevato, il massimo più basso. I diversi passi possono riassumersi nel Figura A. 9.

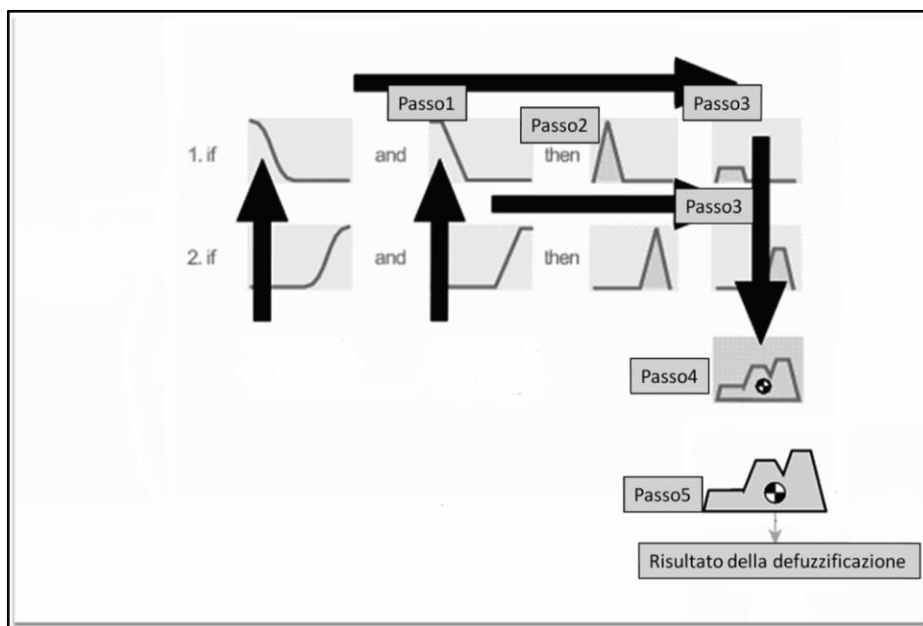


Figura A. 9: Passi del modello fuzzy di Mamdani

Il modello di Takagi-Sugeno-Kang è stato introdotto nel 1985 ed è simile al modello di Mamdani in molti aspetti. I primi due passi del

processo sono uguali; nei successivi la differenza consiste nell'output della funzione MF che è rappresentato da una funzione lineare o costante $z=f(x,y)$; l'output finale sarà un singleton. Questo modello è molto più compatto ed efficiente a livello computazionale e permette di utilizzare tecniche adattive per la definizione della funzione MF ; inoltre è molto adatto per l'analisi matematica, mentre il modello di Mamdani è più adatto alle scienze umane.

Appendice D

La teoria dell'evidenza

Se la teoria della probabilità assume che un individuo conosca l'insieme delle probabilità elementari dell'evento e che derivi da questo tutte le altre mediante unioni, intersezioni, e complementazioni, la teoria dell'evidenza non definisce alcun operatore sull'insieme delle possibilità. Essa lascia spazio alle possibilità, che possono evolversi nel tempo in base al cosiddetto "quadro di discernimento" dell'individuo, che viene detto decisore.

Dato un quadro di discernimento Θ si supponga che il decisore riceva una testimonianza (*body of evidence*) espressa numericamente, che può associarsi ad una serie di possibilità A_1, A_2, \dots, A_N . Ovvero una testimonianza può essere espressa come:

$$\{m(A_1), m(A_2), \dots, m(A_N), m(\Theta)\}$$

con $m(A_i)$ che indica l'evidenza empirica che sostiene la possibilità A_i

Il valore m associato ad Θ non si riferisce a nessuna possibilità in particolare, ma è un'informazione di una scarsa evidenza che sostiene la possibilità A_i . I valori di m vengono normalizzati in tal modo:

$$\sum_{i=1}^N m(A_i) + m(\Theta) = 1.$$

In tal modo i valori della testimonianza per le varie possibilità è pari all'unità.

Si definiscono inoltre due funzioni: la funzione di convincimento e quella della plausibilità. La prima, *belief function* (BF), è la funzione del decisore a favore un'ipotesi H , ovvero le possibili scelte, definita come:

$$Bel(H) = \sum_{A_i \subseteq H} m(A_i) \quad \forall Bel(\emptyset) = 0 \quad Bel(\Theta) = 1$$

La funzione di plausibilità è definita come:

$$Pl(H) = \sum_{H \cap A_i \neq \emptyset} m(A_i) \quad \text{con } Pl(H) = 1 - Bel(\bar{H})$$

$$Bel(H) \leq Pl(H)$$

Queste due funzioni possono essere viste nell'ambito della teoria della probabilità, come l'analogo dell'estremo inferiore e superiore di un intervallo di probabilità.

La teoria dell'evidenza pone in risalto anche dei fatti imprevisi e testimonianze più o meno contraddittorie all'interno di un quadro di discernimento.

Sia quindi $\{m(B_1), m(B_2), \dots, m(B_N), m(\Theta)\}$ una nuova testimonianza che deve essere combinata con $\{m(A_1), m(A_2), \dots, m(A_N), m(\Theta)\}$. E' possibile poter assistere a tre circostanze diverse:

- $A_i \equiv B_i$ e $m(A_i) = m(B_i) \quad \forall i$, ovvero le due testimonianze sono uguali
- $A_i \equiv B_i$ e $\exists i : m(A_i) \neq m(B_i)$, ovvero che due testimoni attribuiscono gradi diversi di affidabilità pur riferendosi alla stessa testimonianza

- \exists almeno un $i : A_i \neq B_j \forall j$ e $m(A_i) \neq m(B_j)$ ovvero che due testimonianze sono diverse e valutano gli stessi eventi in maniera diversa.

Schematicamente si possono rappresentare le due testimonianze come nella Figura A. 10:

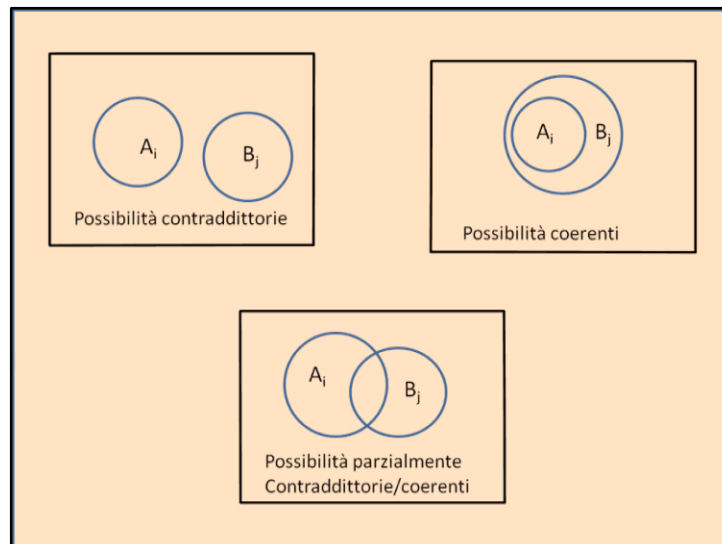


Figura A. 10: Le diverse possibilità di due testimonianze

Se le due testimonianze non hanno nessun elemento in comune, cioè se:

$$m(\Theta) = 0 \quad \forall A_i, B_j : \forall i, \forall j \quad A_i \cap B_j = \emptyset$$

esse saranno considerate come una unica testimonianza e normalizzate. Negli altri casi si utilizza la regola di Dempster-Shafer che permette di costruire una testimonianza $\{m(C_k)\}$ evidenziando gli elementi compatibili tra $\{m(A_i)\}$ e $\{m(B_j)\}$ e tralasciando quelli contraddittori.

Allora, siano date due testimonianze $\{m(A_1), m(A_2), \dots, m(\Theta)\}$ e $\{m(B_1), m(B_2), \dots, m(\Theta)\}$ normalizzate, tale che

$\sum_{i=1}^N m(A_i) + m(\Theta) = 1$ e $\sum_{i=1}^N m(B_i) + m(\Theta) = 1$ devono essere combinate in una testimonianza $\{m(C_1), m(C_2), \dots, m(\Theta)\}$ che soddisfi $\sum_{i=1}^N m(C_i) + m(\Theta) = 1$.

Per ogni possibilità C_k l'ammontare di evidenza empirica combinata sarà:

$$m(C_k) = \frac{\sum_{A_i \cap B_j = C_k} m(A_i) m(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset_k} m(A_i) m(B_j)}$$

Il numeratore rappresenta l'influenza di quegli aspetti della seconda testimonianza che confermano la prima; mentre il denominatore è il complemento a 1 di quegli aspetti della seconda testimonianza che contraddicono la prima. Se non c'è una evidenza contraddittoria il denominatore sarà pari all'unità; aggiungendo un elemento contraddittorio, facendo così diminuire il valore al denominatore, si finisce con il rafforzare il convincimento degli elementi coerenti tra loro e di indebolire quelli in conflitto. La regola può essere applicata solo se le testimonianze sono indipendenti e riferiti alla stessa struttura di discernimento.

Appendice E

Il data clustering

L'attività di clustering, o *cluster analysis* è un processo di organizzazione di oggetti, detti *pattern* in gruppi, detti *cluster*, in cui si riconoscono membri che hanno una forte similarità all'interno del gruppo e una forte dissimilarità tra i patterns degli altri cluster. Gli algoritmi di clustering sono da supporto al *data clustering*, attività del *data mining*. I tipi di dati utilizzati negli algoritmi di clustering possono appartenere alla categoria dei dati quantitativi, categorici o metrici. I primi definiscono per ogni pattern un vettore che ne rappresenta le caratteristiche mediante dati quantitativi; i secondi per i quali sui dati non è possibile eseguire operazioni matematiche o predicati (se non l'uguaglianza); i terzi per i quali è definita una funzione di distanza che associa a due pattern un numero reale che rappresenta la distanza tra essi. Gli algoritmi possono suddividersi in quattro differenti classi: algoritmi partizionanti, gerarchici, basati sulla suddivisione dello spazio e infine basati sulla densità.

Gli algoritmi partizionanti si basano su partizioni del database di partenza ottimizzando la partizione mediante una funzione di similarità, come ad esempio la distanza, in modo che gli elementi che cadono in un cluster siano simili tra loro ma molto differenti rispetto agli altri elementi che appartengono ad altri cluster. Gli algoritmi partizionanti più utilizzati sono il *k-means*, il *k-medoids*, il *CLARA* (Clustering in LARge Applications) e il *CLARANS* (Clustering in LARge Application based upon RANdomized Search). Il *K-means* prevede innanzitutto la scelta arbitraria di un numero di centri iniziali dei cluster, assegna i diversi

elementi ai diversi cluster rispetto alla scelta dei centri,, calcola nuovamente i centri dei cluster che diventano i nuovi centroidi, media pesata dei punti di ogni classe, e prosegue fino al raggiungimento del criterio di iterazione. La Figura A. 11 rappresenta i diversi passi dell'iterazione.

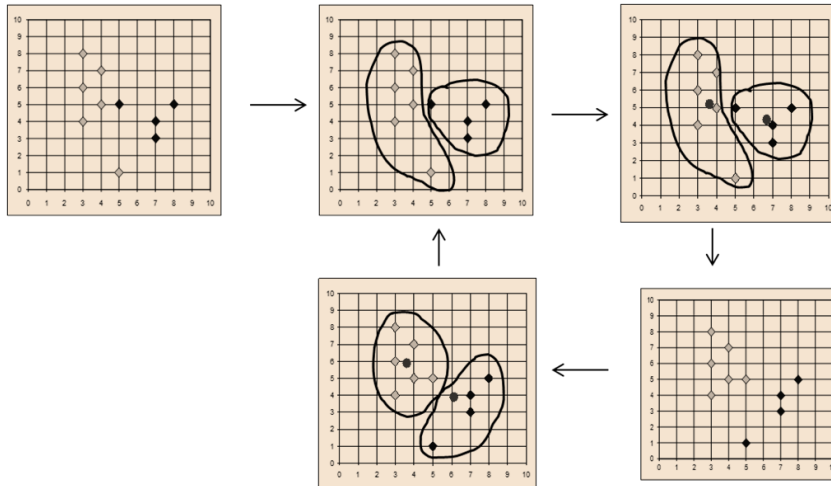


Figura A. 11: Iterazioni del *k-means*

Il *k-menoids* (Figura A. 12) a differenza del *k-means* costruisce il suo centroide tra i punti che hanno la minima distanza media dagli altri punti appartenenti allo stesso cluster, ovvero l'elemento che è localizzato più centralmente (menoide).

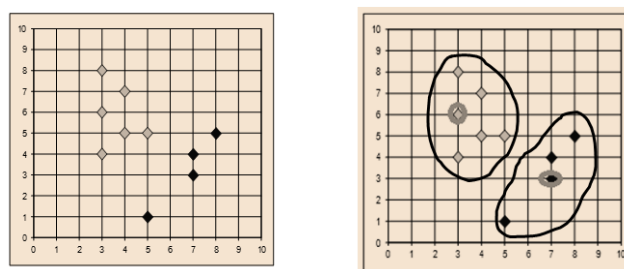


Figura A. 12: I Iterazioni del *k-menoids*

Gli algoritmi gerarchici lavorano raggruppando gli elementi in una struttura ad albero, che suddivide ricorsivamente i dati iniziali. Gli algoritmi agglomerativi utilizzano un approccio *bottom up*, ovvero definiscono ogni cluster contenente un solo elemento e successivamente uniscono più cluster fino ad arrivare ad un unico cluster contenente tutti gli elementi. Quelli detti separativi utilizzano la tecnica detta del *top down* secondo cui si parte con un unico cluster che successivamente viene suddiviso in cluster atomici. La Figura A. 13 rappresenta sinteticamente le due tecniche caratteristiche degli algoritmi AGNES e DIANA.

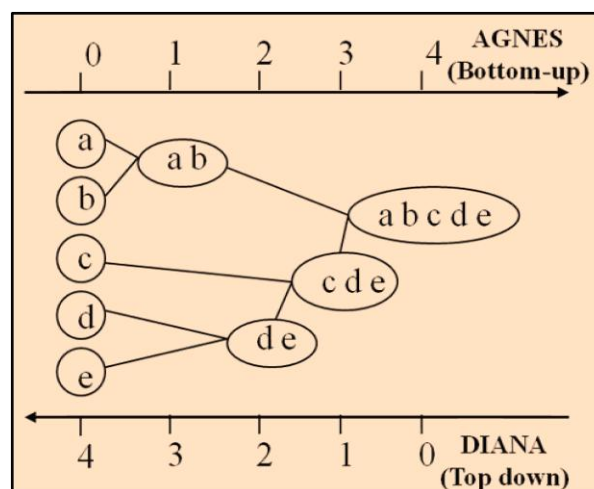


Figura A. 13: Partizionamento di DIANA e AGNES

CHAMELEON è un tipo di algoritmo gerarchico che utilizza un modello dinamico di scomposizione : in una prima fase un algoritmo partizionante per dividere gli elementi in un numero elevato di sub-cluster relativamente piccoli, successivamente utilizza un algoritmo gerarchico agglomerativo e combina con processi iterativi i sub-cluster in cluster identificativi. La Figura A. 14 riporta sinteticamente le fasi dell'algoritmo.

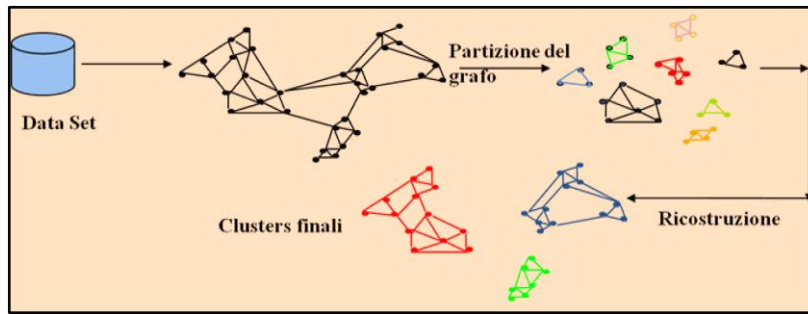


Figura A. 14: Fasi di CHAMELEON

Gli algoritmi basati sulla suddivisione dello spazio (*grid-based*) utilizzano una struttura dati a griglia multirisoluzione: quantizzano lo spazio in un numero finito di celle su cui successivamente vengono effettuate le operazioni di clustering. Le tecniche adottate sono diverse: STING (*Statistical Information Grid approach*) si basa sulla suddivisione del database in tante celle rettangolari che possono corrispondere a diversi livelli di risoluzione. Queste celle formano una struttura gerarchica: ciascuna cella di un livello elevato viene suddivisa per formare un numero di celle ad un livello immediatamente inferiore. L'algoritmo memorizza alcune informazioni statistiche relative agli attributi delle celle della griglia, in cui vengono memorizzati una distribuzione di misura statistica. *WaveCluster* si basa sulla trasformata *wavelet*, CLIQUE (*Clustering In QUEst*) è versatile per database con oltre 20 dimensionalità. Nella Figura A. 15 sono riportati graficamente due approcci di tale tipologia di algoritmo.

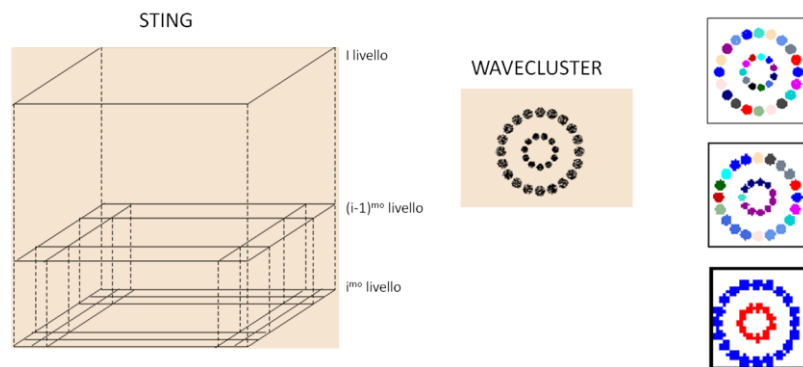


Figura A. 15: alcuni algoritmi basati sulla suddivisione dello spazio

Nella Tab. A 1 sono riportate alcune particolari caratteristiche degli algoritmi menzionati e la velocità computazionale di ognuno di loro.

Tab. A 1: Caratteristiche di alcuni algoritmi

Nome	Caratteristiche	velocità
K-means	Si arresta dopo un certo numero di passi t o quando non si possono più trovare migliori centri k	$O(nkt)$
CLARANS	migliore scalabilità, bisogna definire a priori il numero di cluster; scelta casuale dei centri iniziale, forma ipersferica; outliers creano rumore	$O(n^2)$
CURE	Insieme di centri: cluster non sferici; minore sensibilità al rumore per eliminazione outlier	$O(n^2 \log n)$
CHAMELEON	costruzione di un grafo, individuazione dei cluster piccoli, aggregazione; applicabile anche a dati non metrici; non influenzato da rumore; sensibilità bassa ai parametri di ingresso; rileva cluster di forma arbitraria	$O(n^2)$
BIRCH	Utilizzo di albero bilanciato; migliore scalabilità per metodo di esecuzione dei processi; cluster sferici	$O(n)$
Wavecluster	Spazio suddiviso in celle su cui si effettua il clustering; trasformata wavelet; migliore scalabilità; riconoscimento di cluster non sferici	$O(n)$
DBSCAN	Ogni pattern deve avere nel suo raggio almeno un certo numero di altri pattern; applicata a tutti i dati metrici; identifica gli outliers come rumori	$O(n \log n)$
OPTICS*	Crea un database ordinato sulla struttura di densità dei cluster; successivamente si applica un algoritmo di cluster; è una base per una cluster analysis	$O(n \log n)$

II DBSCAN

Il *Density Based Spatial Clustering of Application with Noise* (DBSCAN) Ester et al. (1996) è un algoritmo ideato per dati spaziali,

secondo cui ogni pattern di un cluster deve avere, in un raggio definito, un certo numero di altri pattern. I pattern a cui si fa riferimento devono essere dati metrici, cioè dati per cui è possibile definire una distanza che rispetti i quattro assiomi della metrica; la distanza viene calcolata tra due punti p e q . I pattern che cadono all'interno del cluster sono detti *core point* mentre quelli ai margini sono detti *border point*; il raggio massimo dell'area in cui si possono trovare i vicini è detto Eps e $MinPts$ è il numero minimo dei pattern. La forma di questa zona intorno al *pattern* dipende dal tipo di funzione di distanza scelta, che viene definita in base al tipo di applicazione. Viene definita una misura \mathcal{E} di vicinanza (*Neighborhood*) di un punto p denotando come simili i punti che hanno una misura di vicinanza $N_{Eps}(p)$ a p i punti che godono della seguente proprietà:

$$N_{Eps}(p) = \{x \in D \mid dist(p, q) \leq Eps\}$$

con D l'insieme di tutti i pattern di tutti i cluster. Questa definizione non basta per definire la densità poiché bisogna definire i punti che sono all'interno del cluster, ovvero i *core point* e quelli che sono ai margini; inoltre, se Eps è troppo piccolo non ci sono risultati, ma se è troppo grande tutti i punti ricadono in un unico cluster. Quindi si definisce un ulteriore parametro $MinPts$ che misura il numero dei punti nel cluster in $N_{Eps}(p)$. Si definiscono i punti esterni (*core point*) come segue:

Un punto p è un *core point* se nel suo intorno di raggio Eps sono presenti almeno $MinPts$ punti:

$$\|N_{Eps}(q)\| \geq MinPts$$

- Un pattern p è direttamente raggiungibile (*directly density-reachable*) da un pattern q relativamente ai parametri Eps e $MinPts$ (raggiungibilità diretta) (Figura A. 16):

1. $p \in N_{Eps}(q)$

2. $N_{Eps}(p) \geq MinPts$ (q è un *core point*)

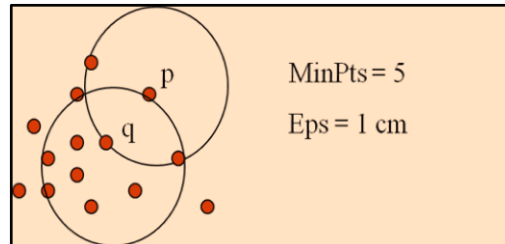


Figura A. 16: Proprietà del DBSCAN

- Un pattern p è detto raggiungibile (*density reachable*) da un pattern q se, assegnati Eps e $MinPts$, esiste una catena di pattern p_1, \dots, p_n , con $p_1 = q$ e $p_n = p$ tale che p_{i+1} sia *directly density-reachable* da p_i (raggiungibilità) (Figura A. 17).

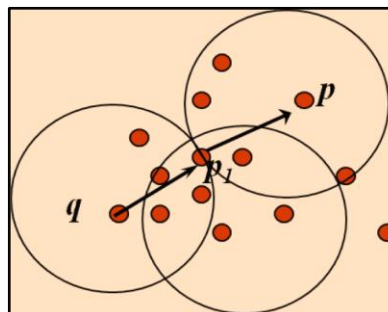


Figura A. 17: Proprietà del DBSCAN

- Un pattern p è connesso (*density connected*) ad un pattern q se esiste un pattern o tale che sia p che q siano *density-reachable* dal

pattern o relativamente agli stessi parametri Eps e $MinPts$ (connessione) (Figura A. 18).

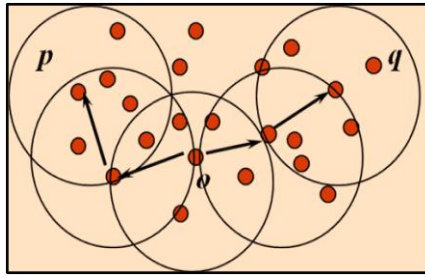


Figura A. 18: Proprietà del DBSCAN

Il DBSCAN si fonda su ulteriori definizioni definite dagli autori:

- Assegnati Eps e $MinPts$ e in database D di un insieme di pattern si definisce un cluster C un sottoinsieme non vuoto di D in cui valgono le seguenti proprietà:

$$\forall p, q: \text{se } p \in C \text{ e } q \text{ è density - reachable da } p \Rightarrow q \in C$$

$$\forall p, q: \text{se } q \in C \text{ e } p \text{ è density - connected da } q \Rightarrow p \in C$$

La prima definizione riguarda la massimalità e la seconda la connettività.

- Viene definito *noise* ogni pattern D che al termine dell'operazione di clustering non risulta assegnato a nessun cluster:

$$\{p \in D \mid \forall i : p \notin C_i\}$$

- *Teorema:* Sia p un pattern di D e $\|N_{Eps}(p)\| \geq MinPts$ allora relativamente ai parametri Eps e $MinPts$ l'insieme $O = \{o \mid o \in D \text{ con } o \text{ density - reachable da } p\}$ è un cluster

- *Teorema:* Assegnati Eps e $MinPts$ sia C un cluster e p un pattern di C tale che $\|N_{Eps}(p)\| \geq MinPts$; allora C coincide con l'insieme O .

Per determinare un cluster DBSCAN si definisce un pattern arbitrario p e si individuano tutti i pattern *density-reachable* assegnati Eps e $MinPts$. Questi due valori dovrebbero essere definiti per ogni cluster, ma nella pratica non è possibile avere questo tipo di informazione; per questo motivo si assegnano dei valori globali, uguali per ogni cluster. Se p è un *core point* si identifica un cluster mentre se è un *border point* si analizza un altro pattern del database. Il database viene analizzato interamente, originando una complessità operativa pari a $O(n)$; inoltre bisogna aggiungere quella dell'interrogazione alla regione specificata, che implica una complessità stimabile di $O(\log n)$. Quindi la complessità globale dell'algoritmo sarà pari a $O(n \log n)$. A differenza degli altri algoritmi partizionanti, il DBSCAN non richiede a priori il numero di cluster da ricercare né dei pattern iniziali per l'inizio della ricerca. I parametri da fissare Eps e $MinPts$ definiscono se un insieme di punti costituisce un cluster o solamente un *noise*, quindi la loro scelta è fondamentale anche se dipende dalle informazioni che si dispongono sul problema da affrontare.

Bibliografia

Bernstein, D., Kornhauser, A. (1996). An introduction to map-matching for personal navigation assistants, *TIDE center*, New Jersey.

Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1), 208-221. Elsevier Science Publishers B. V.

Chou, Y.L., Romeijn, H.E., Smith, R.L. (1998). Approximating shortest paths in large-scale networks with an application to intelligent transportation systems. *INFORMS Journal on Computing* 10(2).163–79.

Chung, E.-H. and Shalaby, A. (2005). A trip bases reconstruction tool for GPS-based personal travel surveys, *Transportation Planning and Technology*, 28 (5) 381–401.

Crawley, M.(2010). *The R Book*. Wiley.

Cui, Y., Ge, S.S. (2003). Autonomous vehicle positioning with GPS in urban canyon environments, *IEEE Transactions on Robotics and Automation* 19 (1), 15–25.

Fu, L., Sun, D., and Rilett, L. R. (2006). Heuristic shortest path algorithms for transportation applications: state of the art. In *Computers & Operations Research*. 33, 3324-3343.

Greenfeld, J. S. (2002). Matching GPS observations to locations on a digital map, paper presented at the 81st Annual Meeting of the *Transportation Research Board*, Washington, D.C., January 2002.

Gustafsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., Nordlund, P., (2002). Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing* 50, 425–435.

Hart, P. E., Nilsson, N. J., Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths, *IEEE Transactions on Systems Science and Cybernetics* (2), pp. 100-107.

Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. Springer NY.

Iacus, S., Masarotto, G. (2006). *Statistica con R*. Seconda edizione. McGraw-Hill.

Jagadeesh, G.R., Srikanthan, T., Quek, K.H. (2002). Heuristic techniques for accelerating hierarchical routing on road networks. *IEEE Transactions on Intelligent Transportation Systems*, 3(4), 301–9.

Joshi R.R. (2001). A new approach to map matching for in-vehicle navigation systems: the rotational variation metric. In *IEEE Intelligent Transportation Systems Conference Proceedings*, 2001, Oakland, CA, USA. pp. 33-38.

Kim., S., Kim, J.-H, Hyun, I.-H. (1998). Development of a map matching algorithm for car navigation system using fuzzy Q-factor algorithm. In proceedings of the *World Congress*

Intelligent Transport System, October, Seoul, Korea.

Kim, W., Jee, G, Lee, J. (2000). Efficient use of digital road map in various positioning for ITS, *IEEE Symposium on Position Location and Navigation*, San Diego, CA.

Kim, S., Kim, J. (2001). Adaptive fuzzy-network based C-measure map matching algorithm for car navigation system, *IEEE Transactions on industrial electronics*, 48 (2), 432-440.

Kumar, K. A., & Rangan, C. P.(2007). Privacy Preserving DBSCAN Algorithm for Clustering. Lecture Notes in *Computer Science*, 4632/2007, 57-68.

Liu, B. (1997). Route finding by using knowledge about the road network. *IEEE Transactions on Systems, Man, and Cybernetics*, 27(4),436–48.

Marchal, F., Hackney, J. K. and Axhausen, K.W. (2005). Efficient map matching of large Global Positioning System data sets: Tests on speed- monitoring experiment in Zürich, *Transportation Research Record*, 1935, 93–100.

Mardia, K.V., Kent, J.T. and Bibby, J.M (2003). *Multivariate Analysis*. Academic Press.

Nicholson.T.A.J. (1966). Finding the shortest route between two points in a network. *Computer Journal*, 9:275–80.

Nielsen, O., Würtz A. and Jørgensen, R. M. (2004). Improved map-matching algorithms for GPS-data - methodology and test on data from the AKTA roadpricing experiment in Copenhagen, 19th European Conference for *ESRI Users*, Copenhagen.

Ochieng, W. Y., Quddus, M. A. and Noland, R. B. (2004). Map matching in complex urban road networks, *Brazilian Journal of Cartography*, 55 (2) 1–18.

Phuyal, B. (2002). Method and use of aggregated dead reckoning sensor and GPS data for map-matching. In: proceedings of the *Institute of Navigation (ION)* annual conference, Portland, OR.

Pyo, J.-S., Shin, D.-H. and Sung, T.-K. (2001). Development of a map matching method using the multiple hypothesis technique, paper presented at *Intelligent Transportation Systems Conference (ITSC)*, Oakland.

Quddus, M.A., Ochieng, W.Y., Zhao, L., Noland, R.B. (2003). A general map matching algorithm for transportation telematics applications. *GPS Solutions* 7 (3),157–167.

Quddus, M. A., Ochieng, W. Y., Zhao, L. and Noland, R. B. (2006a). A high accuracy fuzzy logic-based map matching algorithm for road transport, *Journal of Intelligent Transportation Systems*, 10 (3) 103–115.

Quddus, M.A. (2006b). High Integrity Map Matching Algorithms for Advanced Transport Telematics Applications. Ph.D. Thesis, *Centre for Transport Studies, Imperial College London*, UK.

Quddus, M.A., Ochieng, W.Y. and Noland, R.B. (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions, *Transportation Research C: Emerging Technologies*, 15(5), pp 312 – 328.

Quddus, M.A., Noland, R.B., Ochieng, W.Y. (2009). The effects of navigation sensors and digital map quality on the performance of map matching algorithms, *Geoinformatica*, 13(1), 85-108.

Seebr, G. (2003). *Satellite Geodesy*, De Gruyter, New York

Schüssler, N. and Axhausen, K. W. (2009). Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT), Working Paper, 568, IVT, ETH Zurich, Zurich.

Syed S. and Cannon M.E., (2004). Fuzzy logic based-map matching algorithms for vehicle navigation system in urban canyons, *ION National Technical Meeting*, 26-28 Jan 2004, San Diego, CA.

Velaga, N.R., Quddus, M.A., Bristow, A.L. (2011). Improving the Performance of Topological Map-Matching Algorithm Through Error Detection and Correction. Presented at the annual meeting of the *Transportation Research Board*, 2011, Washington, D.C.

Wagner, D. P. (1997). Lexington area travel data collection test: GPS for personal travel surveys, Final Report, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus.

White, C. E., D. Bernstein and A. L. Kornhauser (2000). Some map matching algorithms for personal navigation assistants, *Transportation Research Part C: Emerging Technologies*, 8 (1–6) 91–108.

Yin H., Wolfson O. (2004). A Weight-based Map Matching Method in Moving Objects Databases; proceedings of the *16th International Conference on SSDBM*.

Yuan, J., Zheng, Y., Zhang, C., Xie, X., Su, G. Z. (2009). An Interactive-Voting-Based Matching Algorithm, In Proc. *MDM*, 2010

Zhang Y., Gao Y. (2008). A Fuzzy Logic Map Matching Algorithm, *Fifth International Conference on Fuzzy System and Knowledge Discovery*, 132-136.

Zhao, Y. (1997). *Vehicle Location and Navigation System*. Artech House, Inc., MA.

Zheng, Y., Quddus, M.A. (2011). Weight-Based Shortest-Path Aided Map-Matching Algorithm for Low-Frequency Positioning Data. Presented at the annual meeting of the *Transportation Research Board*, 2011, Washington, D.C.

Zhou, J., Golledge, R. (2006). A Three-step General Map Matching Method in the GIS Environment: Travel/Transportation Study Perspective, *International Journal of Geographical Information System*, Vol. X, No. X.