



Università degli Studi di Cagliari

## **DOTTORATO DI RICERCA**

**Ingegneria del territorio  
Ciclo XXVII**

# **ANALISI COMPORTAMENTALE DELLA SCELTA DEL PERCORSO ATTRAVERSO L'UTILIZZO DI NUOVE TECNOLOGIE DI ACQUISIZIONE DELLE INFORMAZIONI**

ICAR/05

Presentata da:                      Ing. Alessandro Vacca  
Coordinatore Dottorato:        Prof. Ing. Roberto Deidda  
Tutor/Relatore:                    Prof. Ing. Italo Meloni

Esame finale anno accademico 2013 – 2014



*Desidero ringraziare ed esprimere la mia riconoscenza nei confronti di tutte le persone che, durante questi tre anni di Dottorato, mi sono state vicine e hanno consentito la realizzazione di questo lavoro.*

*Ringrazio in particolare il Prof. Italo Meloni, relatore di questa tesi, per tutti gli insegnamenti ricevuti durante gli anni di collaborazione, per essere stato un costante riferimento e per avermi trasmesso l'entusiasmo ad impegnarmi sempre di più in questo progetto.*

*Vorrei esprimere la mia sincera gratitudine alla Technical University of Denmark e in particolare al Prof. Carlo Giacomo Prato, per l'indispensabile collaborazione nello sviluppo della tesi e per l'amichevole ospitalità durante il mio periodo all'estero.*

*Un sentito ringraziamento va inoltre alla società Systematica s.r.l. di Milano, in particolare all'Arch. Rawad Choubassi e all'Ing. Diego Deponete per essere sempre stati un importante supporto durante tutto il periodo di Dottorato.*

*Ringrazio tutti i colleghi del CRiMM ed in particolare la Sig.ra Catherine Mann per la disponibilità, la pazienza e gentilezza.*

*Ringrazio inoltre tutti i miei amici, vecchi e nuovi, che in diversi modi mi hanno accompagnato durante l'intero periodo di ricerca.*

*Il ringraziamento finale, forse il più importante, va a ai miei genitori, che mi hanno supportato e sopportato durante questi anni di lavoro, spronandomi a raggiungere sempre i miei obiettivi.*



## **ABSTRACT**

In travel demand modeling, route choice is one of the most complex decision-making contexts to understand and mathematically represent for several reasons. Firstly, a large number of available paths may exist between the same origin-destination (OD) pair. Secondly, neither the traveler nor the modeler are aware of all the available alternatives. Thirdly, individual choices are dictated by different constraints and preferences that are difficult to capture by modelers who face increasingly larger datasets where retrieving the exact path chosen by travelers is not always straightforward. Last, there is a lot of uncertainty about travelers' perceptions of route characteristics as well as other characteristics that can influence their choices, such as age, gender, habit, weather conditions and network conditions. This highlights the difficulties encountered for interpreting individual user behavior in greater depth. The rapid advances in GPS devices, has resulted in major benefits for data collection, which now can be recorded automatically and with greater accuracy compared to the techniques used in the past (phone calls, e-mails, face-to-face interviews, laboratory experiments.).

On these basis the main objective of the thesis is then to study route choice using a GPS database. The data were acquired during a survey, named "Casteddu Mobility Styles" (CMS), conducted by the University of Cagliari (Italy) in the metropolitan area of Cagliari between February 2011 and June 2012. Each participant was asked to carry a smartphone with built-in GPS in which an application called "Activity Locator" – implemented by CRiMM (Centre for Research on Mobility and Modeling) – had been installed. A total of 8831 trips were recorded by 109 individuals, of which 4791 referring to the car driver mode. Each GPS track (consisting of a sequence of referenced position points) was then treated with map-matching techniques, through which it was possible to associate each "GPS point" to a link of the network, thus creating the observed route database.

The first objective of the thesis is to understand which are the characteristics of the data acquired during the CMS survey, doing firstly the same analysis that other authors did in their researches based on GPS data. In almost all the previous researches, the GPS data were collected through in-vehicle surveys that make it possible to gather objective information on trips (travel times and distances). Pre-and post-analysis interviews were conducted to gather information about the subjective characteristics of the individuals and GIS platforms were used to study the routes. In the present study, the data were collected using an integrated system able to also record the activities conducted, along with all the characteristics associated thereto. In this way a complete database was created containing all the information (objective and not) concerning the trips. For comparisons with the objectively most convenient paths, then, was used a static macrosimulation model (implemented in CUBE, Citilabs Ltd.) of the entire study area (Cagliari and its metropolitan

area), which reproduces the network characteristics actually encountered by the drivers referring to the data used.

From this first analysis it was observed that when more than one route is taken for repetitive trips between the same OD. In order to understand these particular behavior of users, named also intravariability, discrete choice models were estimated. It's important to note that in the previous GPS-based researches this particular behavior was only identified, without studying it in depth. Several other studies, focused on route switching behavior, tried to understand it applying discrete choice models, but their database were based on data acquired through questionnaires or laboratory experiments, and for the majority the route switching behavior was studied in relation to the trip information provision. The objective of this analysis is then to combine the two fields of the research on route switching, trying to understand it estimating discrete choice models using a GPS based database, closing the gap of the previous researches. The final goal of the model estimations is to understand which are the main attributes of the routes and the characteristics of the users that most influence the choice of an habitual route for the same origin-destination (OD) trip.

After these first analysis, the final objective of the thesis is to apply a route choice model to GPS-based data. Modeling route choice behavior is generally framed as a two-stage process: generation of the alternative routes and modeling of the choice from the generated choice set. The focus of this step of the research is on the bias that might be introduced in the model estimation by the choice set generation process. Specifically, although several explicit choice set generation techniques are found in the literature, the focus is on stochastic route generation and the correction for unequal sampling probability of routes when applying this technique that is easily applicable to large-scale networks. Indeed, stochastic route generation is a case of importance sampling where the selection of the path depends on its own properties, so route choice models based on stochastic route generation must include a sampling correction coefficient that accounts for the different selection probability. In this study is proposed a methodology for calculating and considering this correction factor into MNL-based models with choice sets generated by means of stochastic route generation. Specifically, was decided to look at the sampling correction factor proposed for the random walk algorithm and to calculate the route selection probability in order to exploit this expression. Therefore, a procedure is proposed for the computation of the selection probabilities on the basis of the stochastic generation principle, then the correction factor and last the EPS for model estimation. The modeling analysis confirms the functionality of the proposed approach that has great advantages: (i) it provides insight into the application of stochastic generation in route choice modeling, especially in large-scale networks where the only need is a standard random number generator and a Dijkstra algorithm; (ii) it proposes a simple and manageable procedure from the computational perspective for the

calculation of route selection probabilities and hence the correction factor and EPS for model estimation; (iii) it proves the efficiency of the proposed methodology on revealed preference data in a dense urban network by showing an increase in goodness-of-fit of the model and a shift from illogical to logical sign in parameters estimated for key variables such as travel time.





## Indice

Abstract .....	i
1. Introduzione.....	1
1.1. Inquadramento del fenomeno.....	1
1.2. Inquadramento delle tematiche di ricerca.....	2
1.3. Obiettivi della tesi .....	4
2. Analisi sullo stato dell'arte.....	9
2.1. Acquisizione di dati di percorso tramite sistemi GPS.....	9
2.1.1. Intravariabilità .....	17
2.2. La scelta del percorso.....	22
2.2.1. Panoramica generale sul processo decisionale .....	22
2.2.2. Il processo decisionale nella scelta del percorso .....	25
2.2.3. Punto di vista dell'individuo .....	28
2.2.4. Punto di vista del ricercatore .....	29
2.2.5. La specificazione matematica.....	32
2.3. Generazione dell'insieme di scelta .....	34
2.3.1. Metodi deterministici.....	36
2.3.2. Metodi stocastici.....	41
2.3.3. Metodi ad enumerazione vincolata.....	44
2.3.4. Metodi probabilistici .....	45
2.4. Modelli di scelta del percorso .....	46
2.4.1. Modelli Logit.....	48
2.4.2. Strutture GEV .....	52
2.4.3. Strutture non-GEV .....	55
3. Descrizione dei dati.....	60
3.1. Introduzione .....	60
3.2. Casteddu Mobility Styles.....	61
3.2.1. Selezione 1 - Chi usa Metrocagliari? .....	62
3.2.2. Selezione 2 - Abitudini di viaggio.....	63

3.3.	Campagna di rilievo.....	64
3.3.1.	Week “before” .....	65
3.3.2.	Week “after” .....	65
3.4.	Activity Locator .....	65
3.5.	Dati raccolti.....	69
4.	Trattamento dei dati gps.....	74
4.1.	Il modello dell’area vasta di Cagliari.....	74
4.1.1.	Costruzione del grafo di rete.....	74
4.1.2.	Il modello di domanda .....	78
4.1.3.	Il modello di assegnazione.....	80
4.1.4.	Calibrazione e Validazione del modello del traffico privato.....	81
4.2.	Costruzione del database dei percorsi rilevati.....	84
4.2.1.	Map matching .....	84
4.2.2.	Elaborazione dei dati mappati .....	85
4.3.	Costruzione del database dei percorsi simulati.....	86
4.4.	Analisi dei dati nell’ora di punta del mattino.....	87
4.5.	Intravariabilità.....	93
4.5.1.	Il modello mixed logit .....	95
4.5.2.	Caratteristiche dei dati .....	97
4.5.3.	Caratteristiche delle variabili .....	99
4.5.4.	Risultati .....	99
5.	Stima di un modello di scelta del percorso.....	102
5.1.	Contesto teorico .....	103
5.2.	Metodo .....	104
5.2.1.	Descrizione dei dati.....	108
5.2.2.	Risultati del modello.....	109
6.	Conclusioni.....	115
7.	Bibliografia.....	120



# 1. INTRODUZIONE

## 1.1. Inquadramento del fenomeno

Le problematiche legate alla congestione veicolare, specie a livello urbano, risultano ancora in crescita in tutti i paesi del mondo (con indici più elevati nelle aree urbane dei paesi in via di sviluppo), come pure i suoi effetti negativi. Inoltre è aumentata la sensibilità comune verso l'ambiente e la consapevolezza circa gli effetti del cambiamento climatico dovuti alle emissioni di CO<sub>2</sub> così come il deterioramento della qualità della vita. Ancora, l'attuale periodo di contrazione delle risorse finanziarie limita la realizzazione di interventi infrastrutturali di rilievo nei sistemi di trasporto, che comunque necessitano anche di tempi lunghi di attuazione, e rende indispensabili azioni che mirano a gestire in modo più efficiente la dotazione infrastrutturale esistente. Tra le altre azioni che tentano di superare queste criticità è importante annoverare quelle che, specie nel breve periodo, sono indirizzate a rendere più fluido ed efficiente il traffico urbano, soprattutto automobilistico, che come detto si mantiene sempre su livelli elevati. In questa prospettiva una particolare attenzione dei ricercatori può essere rivolta verso una più profonda conoscenza di ciò che sta alla base delle scelte del percorso degli individui guidatori (utilizzatori del sistema di trasporto), per cercare di indirizzarli, attraverso diverse misure e azioni informative (campagne di informazione, promozione e sistemi avanzati di comunicazione con tecnologie ITS (Intelligent Transport System) e di gestione intelligente del traffico ATMS (Advanced Traffic Management System), politiche di pricing) ad un utilizzo più efficiente ed efficace della rete stradale disponibile. La scelta del percorso è infatti quella dimensione di scelta di viaggio che più si presta ad essere influenzata dall'applicazione di interventi e misure sopra richiamate. La componente comportamentale ha una forte influenza nel meccanismo di scelta, per cui, per esempio se adeguatamente informato e sensibilizzato, l'utente può, semplicemente modificando le proprie scelte di viaggio, decidere di utilizzare le alternative proposte e conseguentemente influire sulle caratteristiche e sul grado di operatività del sistema di trasporto.

Approfondire le conoscenze nell'ambito della scelta del percorso è perciò fondamentale ed inoltre lo è per tutta una serie di altri motivi, tra cui quello relativo alla complessità del processo di decisione che sta alla base dello stesso. Infatti, tra quelli relativi alle scelte di viaggio, questo è uno dei più complessi da interpretare, sia per il numero e il ruolo che differenti aspetti ed attributi svolgono, sia per l'intrinseca dinamicità della scelta stessa. Al momento di effettuare una scelta di viaggio entrano in gioco sia le caratteristiche individuali proprie di chi deve effettuare lo spostamento (sociali, economiche, personali) sia quelle topologiche dell'offerta di trasporto (caratteristiche dei percorsi). Il risultato dell'interazione dinamica di queste due componenti è la scelta che l'utente effettuerà. Rispetto ad altre dimensioni di scelta, poi, questa risente maggiormente degli effetti dell'abitudine, che si

caratterizza per essere ripetitiva e relativamente invariata nel tempo e nello spazio. Diverse ricerche empiriche sul comportamento di scelta del percorso, ma non solo, indicano che gli individui/guidatori sono influenzati da diversi attributi, anche in combinazione, nel ricercare l'alternativa da scegliere: di tempo, di distanza, della congestione, del numero di intersezioni e del tipo di strada, della sicurezza, dell'estetica e del paesaggio, delle condizioni climatiche, delle informazioni disponibili etc.. Inoltre, numerosi altri fattori determinano un comportamento differente tra gli individui, tra cui l'esperienza (Horowitz, 1984; Avineri & Prashker, 2003), le abitudini consolidate (Simon, 1955; Bamberg, et al., 2003), la capacità di apprendimento (Bovy & Stern, 1990; Stern, et al., 1993), il contesto socio-economico, sociale, normativo, la capacità di elaborazione delle informazioni da parte dell'individuo (Spissu, et al., 2011), l'attitudine al rischio in condizioni di incertezza (Kahneman & Tversky, 1979; Katsikopoulos, et al., 2002) ed altre ancora.

La complessità dell'argomento ha fatto sì che, tradizionalmente, l'attenzione dei ricercatori si concentrasse sulle componenti più semplici legate alla scelta, ossia agli attributi oggettivi direttamente misurabili sulla rete come tempi, distanze, costi e, comunque, verso tutto ciò che è legato in maniera più o meno diretta alle caratteristiche operative e topologiche della rete di trasporto. Seguendo questo approccio, si è arrivati a disporre di modelli e software di simulazione che, se da un lato sono risultati adatti a ben rappresentare le condizioni globali di caricamento delle reti, dall'altro non riproducono in maniera adeguata le dinamiche individuali di comportamento (analisi a livello del singolo individuo degli attributi comportamentali) e come tali non possono essere utilizzati per stimare cambiamenti a misure che incidono su caratteristiche proprie dell'individuo. Ciò anche perché i dati sulla base dei quali questi strumenti possono essere calibrati sono costituiti da rilievi aggregati di flussi veicolari su svariate sezioni rappresentative, che forniscono però solo delle informazioni quantitative aggregate sui volumi di traffico. Allo stato attuale, le esigenze del processo decisionale sulle applicazioni dei modelli di trasporto stanno richiedendo un livello sempre più complesso di stima dell'impatto di politiche dei trasporti sempre più di specifiche che vanno al di là della sola e tradizionale espansione della capacità. Da qui emerge che l'aumento della complessità dei modelli e delle loro regole di calibrazione richiede dati di maggiore qualità e dettaglio per comprendere il comportamento di viaggio e le condizioni attuali del sistema di trasporto.

## **1.2. Inquadramento delle tematiche di ricerca**

Recentemente, la ricerca nel settore si è dedicata all'approfondimento della conoscenza degli aspetti più direttamente legati al comportamento, sviluppando diverse teorie aventi l'obiettivo di migliorare l'interpretazione del fenomeno della scelta del percorso, attraverso l'analisi e l'introduzione di fattori propri delle caratteristiche soggettive dell'individuo.

Alcuni autori si sono dedicati all'evoluzione dei modelli di scelta discreta, attraverso la calibrazione di modelli basati su "preferenze rilevate, RP" (Ramming, 2002; Nielsen, 2004; Wolf, et al., 2004; Bekhor, et al., 2006). Altri si sono interessati all'inclusione dei fattori psicologici nella struttura matematica delle forme modellistiche, attraverso l'utilizzo di variabili latenti (McFadden, 2001; Prato & Bekhor, 2006; Prato, 2012). Costituiscono un altro ramo di ricerca i cosiddetti modelli ad apprendimento (learning models), basati sull'ipotesi che le scelte di percorso attuali siano influenzate da quelle effettuate precedentemente, tra cui trovano collocazione i modelli FL (Fudenberg & Levine, 1998) e quelli basati sull'"effetto legge" (law effect) di Thorndike come i modelli REL (Erev, et al., 1999) (dei buoni risultati, ottenuti attraverso l'utilizzo di una particolare strategia, fanno sì che la stessa scelta possa essere effettuata di nuovo). Un altro tema è rappresentato dalle sperimentazioni basate sulla prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), in particolare dell'estensione della stessa, la CPT (cumulative prospect theory) con studi sperimentali mirati a stimare i comportamenti di scelta in funzione di un particolare punto di riferimento, valutando il comportamento (propensione al rischio) degli utenti in condizioni di incertezza (Avineri & Prashker, 2003; Avineri & Bovy, 2008; Bogers, et al., 2005; Xu, et al., 2011); il problema dell'applicazione della CPT è che non esistono ancora studi approfonditi su casi reali.

Una comune esigenza delle ricerche citate è quella di disporre di informazioni affidabili e precise su cui lavorare sugli spostamenti. Infatti, le decisioni di percorso sono molto difficili da interpretare e riprodurre in generale, ancor di più se si aggiunge la mancanza di una base dati robusta, che nella migliore delle situazioni si basa su degli esperimenti simulati, in particolare riguardo alle indagini su preferenze rilevate. In questo campo, la tecnologia GPS è stata identificata come un importante strumento per raccogliere dati di qualità, mostrando di avere diversi vantaggi rispetto ai metodi tradizionali. Per prima cosa, fornisce un'accuratezza e dettaglio migliore sugli attributi spaziali e temporali degli spostamenti rispetto a quanto possono riportare gli utenti e, inoltre, i dati GPS sono stati anche usati per correggere errori di sottostima derivati da indagini "paper and pencil" (PAPI) o telefoniche (Wolf, et al., 2004). Un'altra convenienza è data dalla poca "invasività" sugli utenti del processo di acquisizione dei dati di viaggio, facendo leva su una raccolta dati passiva, e quindi collezionando allo stesso tempo informazioni più dettagliate e più accurate. In aggiunta, da un'ulteriore riduzione del carico sugli utenti attraverso metodi di rilevazione automatica del tipo di attività, luogo, tempo e modo di viaggio, i sistemi di raccolta dati GPS integrati con metodi tradizionali (intervista con successiva telefonata) permettono di porre domande più complesse e precise riguardanti gli spostamenti effettuati. Negli ultimi anni infatti il modo più intrigante per usare i dati raccolti passivamente dal GPS è quello di rimpiazzare i diari di viaggio. Nella modellistica per la pianificazione dei trasporti i dati rilevati con GPS sono stati quindi utilizzati allo scopo di:

- Generare fattori di correzione per i tassi di spostamento;
- Identificare gli schemi di attività;
- Investigare sulle interazioni delle attività in una famiglia e in contesti sociali;
- Identificare i luoghi delle attività;
- Identificare le preferenze di scelta del percorso e scelta modale;
- Esplorare la variabilità della formazione degli schemi di attività e viaggio;
- Identificare le condizioni attuali della rete;
- Valutare il comportamento di ciclisti e pedoni;
- Validare i modelli di domanda;
- Identificare le motivazioni dello spostamento e la tipologia di attività;
- Tracciare l'effettivo percorso utilizzato dall'utente.

### 1.3. Obiettivi della tesi

Il primo obiettivo del presente Dottorato di Ricerca è stato quello di costruire modelli di scelta del percorso che cercano di sfruttare al meglio le opportunità informative messe a disposizione dalle nuove tecnologie GPS. L'ultima generazione delle indagini svolte sfruttando questa tecnologia è quella che prevede una prima raccolta delle informazioni di base sull'individuo e la famiglia, ed una fase successiva in cui i "data loggers" GPS (implementanti particolari algoritmi e modelli che permettono di generare tutte le informazioni utili sullo spostamento) sono usati dai partecipanti all'indagine per tracciare i loro movimenti. Questa combinazione di dati spazio-temporali ad alta risoluzione, unitamente al minor onere da parte degli utenti, consentendo la raccolta su più giorni, permette di investigare più in dettaglio su certi aspetti del comportamento di viaggio tra cui la variabilità nella scelta del luogo dove si compie l'attività, nella scelta del modo e nella scelta del percorso. Inoltre, questi dati possono supportare la riduzione delle dimensioni del campione, compensando in qualche modo alcuni, se non tutti, costi addizionali legati ai sistemi di raccolta dati GPS (Stopher, et al., 2008).

È sulla base di queste considerazioni che l'Università di Cagliari ha promosso la campagna d'indagine "Casteddu Mobility Styles" (CMS), condotta tra febbraio 2011 e giugno 2012 nell'area metropolitana di Cagliari (Italia). Lo scopo principale era quello di analizzare il comportamento di viaggio degli utenti del sistema di trasporto per: 1) identificare quali siano i fattori in grado di generare, nei confronti di utenti utilizzatori dell'automobile, dei cambiamenti volontari di comportamento di viaggio, in particolare per ciò che riguarda la scelta del modo di trasporto con il quale effettuare i propri spostamenti quotidiani; 2) valutare l'efficacia di una cosiddetta "soft measure" (in particolare un programma personalizzato di viaggio) 3) definire e stimare modelli di domanda che includessero effetti

dinamici della partecipazione alle attività/viaggi ed effetti di eterogeneità inter-periodale e inter-personale nelle scelte di viaggio.

Nell'ambito di questa attività di ricerca è stato possibile acquisire, attraverso un'applicazione mobile chiamata Activity Locator (Meloni, et al., 2011) e installabile in uno smartphone con GPS integrato, le attività svolte, gli spostamenti e i relativi tracciati di 109 utenti per 14 giorni ciascuno, 24/24h. Per le prime sono risultate disponibili il luogo di svolgimento, l'orario di inizio e di fine (e quindi la durata), la compagnia e il tipo di attività stessa, mentre per ciò che riguarda gli spostamenti (ne sono stati registrati 8831 complessivi), grazie al dispositivo GPS, è stato possibile ottenere, oltreché il tracciato spaziale, anche le informazioni temporali ad esso legate come l'orario di inizio e fine. Il database di dati utili per le analisi di scelta del percorso consiste in 4791 spostamenti nella modalità auto-guidatore.

Una prima fase di questo lavoro è consistita nell'analizzare, anche in confronto con altre ricerche presenti in letteratura, la differenza tra i percorsi reali osservati con quelli di minimo costo, calcolati secondo la minimizzazione di alcuni attributi (come tempi e distanze di percorrenza), che vengono utilizzati nelle procedure di modellizzazione dei flussi di traffico (assegnazione).

Un elemento che distingue il presente studio da altri presenti in letteratura è che per i confronti tra i percorsi osservati e quelli simulati non è stata usata una piattaforma GIS, come nella maggior parte degli studi precedenti, bensì un modello di macrosimulazione calibrato per il contesto di analisi. Quest'ultimo è stato implementato in CUBE (Citilabs), e consente di avere una base di confronto solida e rappresentativa delle condizioni di traffico sulla rete diversamente da quanto avverrebbe utilizzando un GIS, nel quale si potrebbero fare solamente delle macro assunzioni generali, perdendo quindi accuratezza e realismo dei risultati. L'applicazione è stata effettuata su un sottocampione di CMS, relativo a tutti gli spostamenti avvenuti in modalità auto-guidatore nell'ora di punta del mattino, congruente quindi con le condizioni di traffico simulate con il modello.

Questa prima analisi ha evidenziato un interessante fenomeno legato all'intravariabilità dei percorsi effettuati da uno stesso utente che, nel caso di spostamenti tra la stessa coppia origine-destinazione, spesso non utilizza lo stesso itinerario. Questo è un interessante tema di ricerca, perchè le scelte di viaggio (in particolare la scelta del percorso) sono fortemente influenzate dall'abitudine, specie sulle O/D casa lavoro, che costituiscono uno spostamento per il quale l'utente ha la possibilità di acquisire molta esperienza che gli permette di sviluppare una buona conoscenza della porzione di rete utilizzata. L'analisi della letteratura a riguardo ha evidenziato una lacuna importante in questa direzione. Infatti, tale fenomeno è stato affrontato secondo due diversi punti di vista: uno sperimentale, che, attraverso l'analisi di rilievi eseguiti con i GPS, si sono limitati semplicemente a notare la presenza di questo



comportamento senza cercare di approfondirne il perchè; l'altro simulativo, che attraverso la costruzione di modelli matematici hanno indagato sulle dinamiche alla base del cambio di percorso (route switch), valutando gli effetti derivati dalla somministrazione di informazioni prima/durante e dopo il viaggio, in particolare utilizzando dati su preferenze rilevate o dichiarate raccolti tramite dei questionari oppure derivanti da esperimenti effettuati in laboratorio.

Un obiettivo specifico del percorso di ricerca ha riguardato quindi l'approfondimento di questa tematica. Disponendo di informazioni dettagliate sui percorsi realmente utilizzati (due settimane di rilievo consecutive per ciascun utente) si sono estratte le informazioni sulla scelta di percorso, per uno stesso utente, tra la stessa coppia origine/destinazione. L'analisi di questi dati ha mostrato come nell'ora di punta del mattino circa il 29% degli utenti non usi sempre lo stesso percorso nello spostamento casa-lavoro. Per analizzare in profondità questo fenomeno dell'intravariabilità si è proceduto ad applicare un modello di scelta discreta al database di dati GPS a disposizione. Nel dettaglio, le analisi sono state effettuate su un sottocampione degli spostamenti avvenuti nell'ora di punta del mattino, con la caratteristica di essere ripetuti tra la stessa coppia origine-destinazione. Sono state esaminate le caratteristiche del percorso, dello spostamento e dell'individuo che hanno maggiore influenza nel determinare che un itinerario possa essere scelto abitualmente tra la stessa coppia origine e destinazione.

L'ultima fase della ricerca effettuata durante il triennio del dottorato, sei mesi dei quali sono stati trascorsi presso la Technical University of Denmark come visiting student, ha avuto come obiettivo la costruzione di un modello di scelta del percorso ed, all'interno di questo, sulle tecniche di generazione dell'insieme di scelta, che nel caso specifico sono state definite utilizzando dati reali, quelli rilevati attraverso il GPS.

Il tema della generazione dell'insieme di scelta è di particolare interesse sia nel mondo accademico che applicativo, perchè la sua non corretta risoluzione può comportare errori nel processo di stima dei modelli. Un modello di scelta del percorso che ignori il problema della generazione dell'insieme di scelta, assumendo che ogni individuo consideri quello universale contenente tutte le alternative possibili, potrebbe risultare seriamente mis-specificato (Ben Akiva & Boccara, 1995). Una corretta rappresentazione dell'insieme di scelta gioca quindi un ruolo chiave, come hanno evidenziato numerosi ricercatori (Stopher, 1980; Williams & Ortuzar, 1982; Başar & Bhat, 2004; Swait & Ben-Akiva, 1987). In particolare, nel presente lavoro di ricerca si è posta l'attenzione sull'errore che potrebbe introdursi nelle stime dei modelli in seguito al processo di generazione dell'insieme di scelta.

Nel dettaglio, nonostante sia possibile trovare in letteratura diverse tecniche di generazione esplicite dell'insieme di scelta (per una revisione generale consultare Prato, 2009), il lavoro si focalizza su quella stocastica, che è quella più pratica da usare soprattutto in casi di reti

reali, perché i generatori di numeri casuali sono facilmente utilizzabili in relazione all'ambiente operativo dove è effettuata la modellazione della scelta di percorso. La generazione stocastica consiste nella ripetuta ricerca del percorso di costo minimo, dove il calcolo di quest'ultimo avviene attraverso l'estrazione casuale delle impedenze (tempi, costi, etc.) degli archi da determinate distribuzioni di probabilità. Le assunzioni principali alla base di questo metodo sono legate alla selezione della tipologia di distribuzione di probabilità da utilizzare, perché la qualità e il numero dei percorsi estratti dipende dalla definizione dei parametri con cui è identificata. L'obiettivo che ci si pone inizialmente è quindi quello di identificare i parametri ideali per la costruzione di un insieme di scelta utile. Si utilizzerà una distribuzione gamma perché impedisce l'estrazione di costi negativi, testando diversi valori della varianza in modo da trovare quelli che permettono di riprodurre al meglio i percorsi osservati. Se i parametri della distribuzione di probabilità e il numero di estrazioni sono scelti correttamente, la generazione stocastica ha il vantaggio di generare un gran numero di percorsi attrattivi unici che ben riproducono le osservazioni. Tuttavia, questa tecnica è un caso, noto in letteratura, di "importance sampling", perché la probabilità di selezione di un percorso dipende dalle proprietà dello stesso, come la lunghezza o il tempo di percorrenza, e potrebbe anche dipendere dalla sovrapposizione spaziale dei percorsi. Quindi, i modelli di scelta del percorso basati sulla generazione stocastica dell'insieme di scelta hanno bisogno di un fattore di correzione per tenere conto della diversa probabilità di selezione delle alternative ed evitare una errata stima dei parametri (Frejinger, 2007; Frejinger, et al., 2009; Bovy, et al., 2009). Siccome, nel caso di applicazione della generazione stocastica, in letteratura non è stato ancora definito nessun correttore di questo tipo, l'obiettivo della ricerca è quello di definire una metodologia di correzione appropriata per l'utilizzo di questa tecnica, dato l'interessante potenziale di applicabilità.

L'attenzione è ricaduta allora su un fattore di correzione definito recentemente per l'algoritmo di random walk (Frejinger, et al., 2009), che dipende dalle probabilità di selezione associate ai percorsi dell'insieme di scelta (che sono parte integrante del procedimento di calcolo e quindi immediatamente ricavabili) e dal numero di volte che uno stesso percorso è estratto durante il processo di generazione. Tuttavia, il random walk non viene spesso utilizzato nelle applicazioni pratiche a causa della natura combinatoria dello stesso, specialmente quando si lavora su dati ad alta risoluzione, che si traduce in elevati tempi computazionali (Kaplan & Prato, 2012; Rieser-Schüssler, et al., 2013). Si è deciso allora di utilizzare il fattore di correzione già specificato per l'applicazione della tecnica di generazione stocastica. Allo scopo, occorre quindi avere a disposizione le informazioni con cui è definito: numero di estrazioni dello stesso percorso e probabilità di selezione dello stesso. Il primo dato è facilmente ricavabile, mentre allo stato attuale non esiste una metodologia per il calcolo delle probabilità di selezione nel caso di applicazione della generazione stocastica. Nella presente ricerca si propone quindi una metodologia per il

calcolo di queste ultime, utilizzando come riferimento la teoria del percorso di costo minimo, che è inoltre una componente principale della struttura teorica alla base della stessa tecnica di generazione.

L'applicazione del metodo proposto è avvenuta sull'intero database di CMS in modalità auto-guidatore, per un totale di 3656 spostamenti utilizzabili. Sono stati stimati diversi modelli per identificare i migliori valori delle probabilità di selezione, confrontando inoltre i modelli stimati con e senza correzione. I risultati mostrano che il fattore di correzione migliora l'adattamento del modello e, ancora più importante, corregge parametri con segni illogici presenti nelle stime dei modelli senza correzione. Il contributo di questa parte della ricerca è costituito quindi nell'apertura di nuovi orizzonti per l'applicazione della generazione stocastica nella modellazione della scelta del percorso, perché questa tecnica è la più efficiente per generare gli insiemi di scelta; proporre una procedura semplice e gestibile dal punto di vista computazionale per il calcolo delle probabilità di selezione e quindi per il fattore di correzione per la stima del modello; provare l'efficienza della metodologia proposta su dati rilevati in una rete urbana densa.

Il proseguo della tesi è strutturato nel modo seguente: il secondo capitolo riporta la revisione bibliografica, che spazia dagli studi sui dati GPS a quelli in cui è modellata l'intravariabilità, unitamente all'intera paronimica sulla scelta del percorso, presentando assunzioni alla base, specificazione matematica, modelli esistenti; il terzo capitolo riguarda la descrizione della campagna d'indagine Casteddu Mobility Styles, dettagliando in che modo sono stati raccolti i dati e che tipo di informazioni si hanno a disposizione; il quarto capitolo riporta le analisi effettuate sui dati GPS riferiti agli spostamenti nell'ora di punta del mattino, unitamente a quelle effettuate per studiare il fenomeno dell'intravariabilità; il quinto capitolo presenta la metodologia per la derivazione del fattore di correzione per il campionamento e il calcolo delle probabilità di selezione dei percorsi, così come l'applicazione sul database di CMS; il sesto capitolo riporta le conclusioni.

## **2. ANALISI SULLO STATO DELL'ARTE**

### **2.1. Acquisizione di dati di percorso tramite sistemi GPS**

Come evidenziato nel capitolo precedente, l'incremento della disponibilità di dati di viaggio acquisiti tramite tecnologie basate sul posizionamento, come i dispositivi GPS, combinata con la disponibilità di interfacce applicative di programmazione aperte (APIs) e software open source (OSS), ha catturato l'interesse nell'applicazione di dati GPS per le previsioni di viaggio, analisi di pianificazione, e gestione dei sistemi di trasporto. Frequentemente, tuttavia, l'attrazione iniziale da parte delle agenzie pubbliche su questi dati di viaggio così dettagliati si è scontrata con gli ostacoli dovuti al costo, problemi con l'integrazione in paradigmi modellistici esistenti, tematiche legate alla privacy, errori di campionamento e difficoltà nella gestione dei dati. Il GPS o altre tecnologie di tracciamento possono fornire una profonda comprensione dei comportamenti di viaggio e dei modelli di attività che superano le esigenze modellistiche tradizionali (come tempi di viaggio e numero di spostamenti) e che integrano misure standard per le performance del sistema (ad esempio per le velocità o la congestione) (Wolf, et al., 2014).

In passato, i sistemi utilizzati per il reperimento delle informazioni sulla scelta del percorso erano basati su delle interviste effettuate ai viaggiatori, effettuate attraverso telefonate, mail o interviste dirette, nelle quali gli utenti fornivano le informazioni relative al percorso seguito (Ben-Akiva, et al., 1984; Mahmassani, et al., 1993; Ramming, 2002; Prato, et al., 2005). Essendo condizionati dalla memoria degli intervistati, i dati spesso risultano imprecisi e poco attendibili (Murakami & Wagner, 1999; Barbeau, et al., 2009). La diffusione dei sistemi informatici, unita al rapido progredire dei moderni dispositivi GPS, ha portato dei benefici importanti per la raccolta dei dati, i quali ora si possono registrare in maniera automatica, in formato elettronico e con, soprattutto, una più alta precisione rispetto a quelli raccolti attraverso le interviste individuali (Murakami & Wagner, 1999; Nakazato, et al., 2006; Hato, et al., 2006; Bricka, et al., 2009; Barbeau, et al., 2009). Tuttavia, la qualità dei dati GPS è vincolata sia dall'efficienza tecnologica dello strumento e del sistema satellitare sia dal contesto in cui si svolge l'indagine; inoltre i dati raccolti sotto forma di coordinate devono essere sottoposti ad un processo di "map-matching", ossia allocare ciascun punto GPS al corrispondente arco di appartenenza nella rete (Christopher, et al., 2000; Zhou & Golledge, 2006; Quddus, et al., 2007; Marchal, et al., 2005; Schuessler & Axhausen, 2009; Fafieanie, 2009).

Negli ultimi dieci anni, nonostante i limiti sopra descritti, i dati raccolti attraverso strumenti GPS o sistemi integrati più evoluti (PDA; smartphone, etc.) sono stati utilizzati per acquisire informazioni sia spaziali che temporali sugli spostamenti. In particolare, sono stati utilizzati nell'analisi dei comportamenti di scelta di attività e viaggio (Asakura & Iryo, 2007; Stopher &

Greaves, 2007), per approfondire i processi dinamici di programmazione e riprogrammazione delle attività (Zhou & Golledge, 2007) per migliorare la comprensione dei comportamenti spaziali (Schönfelder & Axhausen, 2004), per quantificare la produzione di emissioni da trasporto privato (Beckx, et al., 2010), per calibrare i processi di generazione dell'insieme dei percorsi utilizzabili (Fafieanie, 2009) ed infine misurare l'effetto delle politiche commerciali di cap and trade sugli schemi di attività personali (Meloni, et al., 2011). Il GPS è stato inoltre utilizzato anche per cercare di migliorare la conoscenza degli attributi che governano la scelta del percorso, così come analizzato da diversi autori (Jan, et al., 2000; Li, et al., 2005; Parkany, et al., 2006; Papinski, et al., 2009; Papinski & Scott, 2010; Zhu & Levinson, 2010; Spissu, et al., 2011; Arifin & Axhausen, 2011).

In particolare, Jan, et al. (2000) si pongono l'obiettivo di comprendere come i dati oggettivi sulla scelta del percorso possano aiutare a capire le differenze tra il comportamento reale (acquisito con un dispositivo GPS di tipo "in-vehicle") e quelle interpretate attraverso modelli di assegnazione del traffico sulla rete, sia dal punto di vista pratico che teorico. Allo scopo, utilizzano un sotto-campione di un database di circa 3000 spostamenti ottenuti in una settimana d'indagine che ha riguardato 100 famiglie (216 guidatori) della città di Lexington, Kentucky. Le informazioni a disposizione sono relative ad attributi della rete come distanze, velocità e capacità, mentre per ogni percorso rilevato (contraddistinto da un ID relativo all'utente) si conoscono tempi e distanze ricavate dalla loro implementazione in GIS. Il database informativo non contiene nessuna informazione relativa a semaforizzazioni o ritardi per le manovre di svolta. Vengono calcolati (a) tempi oggettivi di percorrenza della rete, ipotizzando condizioni di deflusso pari all'applicazione di una riduzione (circa il 15%) alla velocità a flusso libero; (b) tempi di percorrenza misurati attraverso il GPS; (c) il tempo del percorso più corto tra la stessa OD, ottenuto nello stesso modo di (a). Da ciò vengono calcolate le percentuali di archi condivisi tra due percorsi con la stessa OD, trovando che per la maggior parte delle coppie vi è una percentuale di condivisione di circa il 60% o poco più. Nello stesso studio, Jan, et al. (2000) introducono il PDI (path deviation index), come rapporto tra l'area sottesa tra i due percorsi e la lunghezza del tragitto più corto, per determinare la distanza spaziale tra due percorsi che, nel caso fosse particolarmente contenuta, consentirebbe di ritenere uguali i percorsi tra le coppie con un basso PDI. Osservano inoltre che gli utenti abitualmente seguono lo stesso percorso per lo stesso spostamento, con il PDI che cresce all'aumentare della distanza tra le OD (distance deviation index), mentre il tempo di percorrenza rimane simile a quello dei percorsi più brevi. Lo stesso paper osserva che le informazioni raccolte attraverso il GPS erano limitate e confida nello sviluppo delle strumentazioni a disposizione, che dovranno essere in grado di raccogliere più tipologie di dati.

Li, et al. (2005) sostengono che la comprensione del comportamento di scelta del percorso analizzata attraverso assunzioni una base dati più realistica possa aiutare a migliorare gli algoritmi matematici di assegnazione. A questo proposito nel loro studio vengono utilizzati i dati rilevati in un'indagine GPS in-vehicle, conosciuta come "Commute Atlanta", dei quali scelgono un sottocampione riferito ad un periodo di 10 giorni, relativo a 182 guidatori appartenenti a 138 famiglie, delle quali si conoscono le caratteristiche socio-economiche. Di ogni percorso registrato sono noti gli orari di partenza e arrivo, tempo di percorrenza e la distanza. Considerando gli spostamenti avvenuti in giorni feriali nell'intervallo orario tra le 5:00 e le 10:00 del mattino, trovano che circa il 40% dei commuters (pendolari) usa solo un percorso per il proprio spostamento, mentre il 60% ne usa almeno due. Considerando come "routine" l'utilizzo dello stesso percorso almeno due volte durante il periodo di studio, si ha che circa il 33% utilizza dei percorsi-routine, valore maggiore rispetto a quanto trovato da Abdel-Aty (Abdel-Aty, et al., 1994), che era pari al 15,5%. Li, et al. (2005) sviluppano inoltre un modello logit binario riferito alle scelte di percorso, utilizzando come variabile dipendente l'utilizzo di uno o più percorsi tra la stessa OD. Nella specificazione del modello vengono introdotte le caratteristiche socio-economiche individuali e della famiglia, le informazioni sugli spostamenti, le caratteristiche del percorso più utilizzato. I risultati mostrano il forte potere esplicativo, sulla variabile dipendente di scegliere più percorsi, della flessibilità dell'orario di lavoro e della catena degli spostamenti rispetto alle caratteristiche socio-demografiche dei pendolari e i relativi attributi di percorso. Nel paper viene osservato che uno studio di questo genere possa acquisire maggiore precisione se si avessero a disposizione dettagliate informazioni circa le attività svolte e gli spostamenti effettuati per periodi di tempo lunghi.

Per Parkany, et al. (2006) l'obiettivo della ricerca è quello di investigare sulla diversione delle scelte di percorso rispetto a quella basata sul minimo tempo, con riferimento all'intervallo orario di picco o di morbida e tra i dati rilevati e dichiarati. Nello studio vengono utilizzati dei dati raccolti tra marzo 2002 e luglio 2003 a Lexington (Kentucky), dove un ricevitore GPS è stato installato in un veicolo per ognuna delle 276 famiglie partecipanti all'indagine, che vengono monitorate per un periodo di 10 giorni. Lo studio prevedeva anche una seconda indagine, seguente a quella realizzata con la raccolta dati GPS, attraverso l'invio tramite mail di un questionario, fattore che ha determinato l'utilizzo di un sottocampione di 106 guidatori (per avere una buona corrispondenza con la mailback interview). Nella post-intervista sono stati raccolti dati sulle abitudini di viaggio, preferenze di percorso tra ora di picco e di morbida, fattori relativi agli atteggiamenti (attitudes). Per calcolare la diversione viene utilizzato il rapporto tra il tempo di percorrenza del percorso scelto rispetto a quello più corto (time deviation index), indicando come "diverters" gli utenti per i quali il rapporto supera il 10%. Gli autori trovano che le ragioni della diversione stanno nel fatto che, nelle scelte, contano altri fattori oltreché il tempo e che gli utenti potrebbero non avere abbastanza

informazioni e abilità per minimizzare accuratamente il tempo di viaggio, confermando la teoria che tra una stessa OD sia usato più di un percorso. Nello studio vengono stimati inoltre alcuni modelli per valutare la consistenza dei percorsi indicati nelle preferenze dichiarate e la differenza riscontrata con quelle rilevate, mostrando come variabili tipo abitudini di viaggio, atteggiamenti e caratteristiche del percorso siano direttamente legate alla scelta del percorso.

Papinski, et al. (2009) si interrogano su come gli utenti pianifichino il proprio spostamento e su quale sia la sequenza di pianificazione. I dati utilizzati si riferiscono ad un campione di 31 utenti, dal quale selezionano 21 percorsi home-work, registrati con il GPS, al fine di valutare la deviazione con i tragitti pianificati (essendo il percorso casa-lavoro un'attività programmata, si presuppone che lo stesso sia pianificato in anticipo). L'indagine si articola in tre fasi: una pre-intervista, la raccolta dati e una post-intervista. Gli autori identificano 4 tipologie di pianificazione del percorso: prima le strade principali, in sequenza, mista, nessuna pianificazione. Mostrano inoltre come il 20% dei partecipanti faccia delle deviazioni rispetto al percorso pianificato (per varie ragioni, che vanno dalla voglia di sperimentare un altro percorso al cambio di umore). In termini spaziali, trovano che i percorsi brevi si prestano a maggiori deviazioni, per il fatto che gli utenti hanno una maggiore conoscenza delle strade locali.

Papinski & Scott (2010) sostengono che per supportare al massimo l'analisi di scelta del percorso occorrono degli strumenti in grado di processare i dati a disposizione in modo efficiente e con poche operazioni. Allo scopo, illustrano lo sviluppo e l'applicazione di un kit di analisi di scelta del percorso (route choice analysis, RCA) in grado di calcolare automaticamente più di 40 variabili relative alle caratteristiche di percorso, che consentono di confrontare gli itinerari osservati con quelli più brevi (calcolati attraverso un apposito modulo integrato nel RCA). Nella loro ricerca utilizzano dei dati osservati col GPS relativi ad un campione di 237 guidatori, partecipanti ad un'indagine di due giorni (per spostamenti di tipo casa-lavoro), strutturata in tre fasi: 1) compilazione un diario cartaceo, 2) registrazione (due giorni) con il GPS, 3) intervista post-indagine. Ogni utente, nel periodo di registrazione, veniva equipaggiato con un palmare il quale integrava la strumentazione GPS necessaria. Nei casi in cui si riscontrava un conflitto tra quanto dichiarato nel diario e i dati GPS, o nei casi in cui mancasse qualche dato, il partecipante veniva contattato per poter completare l'informazione. Attraverso il RCA gli autori calcolano i percorsi più brevi con funzioni di costo basate sulla distanza e sul tempo per poi confrontarli, per mezzo delle variabili che il kit è in grado di calcolare, con quelli realmente osservati; calcolano inoltre il RDI, indice di direzionalità, che misura il grado di "eccesso di lunghezza" del percorso. Nello studio vengono osservate differenze significative tra i percorsi reali e quelli simulati, soprattutto in termini di tempo, distanza, velocità, utilizzo di alcune tipologie di strade e numero di svolte.

Considerando i percorsi più brevi in termini di distanza, trovano che questi sono costituiti da circa il 34% di strade a scorrimento veloce e autostrade, contro il 44% degli osservati che, inoltre, presentano valori minori per il numero di svolte. I percorsi più brevi hanno inoltre un RDI di 1.38 e 1.34 (i decimali rappresentano la % di lunghezza in più rispetto alla distanza in linea retta tra l'origine e la destinazione considerata), contro l'1.53 degli osservati. Concludono sostenendo che le funzioni di costo basate esclusivamente sul tempo o sulla distanza non sono in grado di riprodurre adeguatamente le dinamiche degli spostamenti reali.

Levinson & Zhu (2011) vogliono esplorare il fenomeno che, in contrasto con la teoria dell'equilibrio dell'utente, fa sì che gli individui, in relazione a spostamenti ripetuti tra la stessa OD, scelgano più di un solo percorso. Usano dei dati raccolti tramite un GPS installato direttamente sul veicolo, per un campione di 657 spostamenti home-work avvenuti nell'area di Minneapolis, Saint Paul, Minnesota (Twin Cities), relativi a 95 utenti. I dati registrati sono stati elaborati al fine di poter essere mappati su una rete implementata in GIS ed essere quindi analizzati. Impostando come diversi i percorsi che si discostano almeno del 30% in termini di distanza, trovano che circa il 40% degli utenti monitorati utilizza più di un percorso. Confrontano poi i tempi medi di percorso simulati e la deviazione standard con quelli rilevati dal GPS, mostrano come il loro modello, nella maggior parte dei casi, sottostimi il tempo (probabilmente a causa del fatto che il modello da loro utilizzato riproduca in modo poco efficace i ritardi alle intersezioni che, invece, il GPS registra correttamente). Concludono affermando che tendenzialmente l'utente tende a minimizzare il tempo di viaggio, rimanendo tuttavia entro una certa soglia di "affidabilità" del percorso.

Spissu, et al. (2011) sviluppano una metodologia che permette di analizzare dei dati GPS riferiti a percorsi realmente osservati, con lo scopo di valutare il grado di intra/intervariabilità tra gli individui e per confrontare gli itinerari realmente scelti rispetto a quelli di minimo costo. Utilizzano un campione costituito da 26 studenti dell'Università di Cagliari, monitorati per un periodo di due settimane attraverso un sistema di indagine personale chiamato "Activity Locator", con la possibilità di registrare, oltre che i tracciati dei percorsi, anche le diverse attività che gli utenti hanno compiuto durante le giornate di rilievo. Trovano che uno stesso individuo, tra una specifica coppia OD, utilizzi per circa il 92% un solo percorso (intravariabilità); estendendo l'analisi a più utenti (per spostamenti tra la medesima coppia OD), trovano che un solo percorso è utilizzato per circa il 72% dei casi, due per il 24% e il restante per tre o più (intervariabilità). Calcolano poi l'indice di sovrapposizione (overlapping index), relativo alla percentuale di archi comuni tra il percorso reale e quello simulato, riscontrando che circa il 70% delle registrazioni abbia dei valori compresi tra 0 e 0.6, mentre meno del 30% presenta dei valori maggiori e quindi caratteristiche più simili al rispettivo itinerario di minimo costo. Valutano poi il time



deviation index e il distance deviation index (variazioni percentuali del tempo o della distanza del percorso osservato rispetto a quello simulato), trovando una tendenza generale degli utenti nel minimizzare la distanza per gli spostamenti brevi (circa 4 km), mentre il tempo viene tenuto maggiormente in considerazione per itinerari più lunghi (più di 7 km).

Arifin & Axhausen (2011) riportano i risultati di un'analisi riguardante il comportamento dinamico di scelta del percorso e del modo per 93 utenti che effettuano degli spostamenti sistematici a Jakarta. I dati sono stati raccolti attraverso questionari ed un sistema GPS per un periodo di rilevamento pari ad una settimana, nel periodo tra luglio e settembre 2010. Attraverso un apposito algoritmo elaborano le informazioni registrate identificando origine, destinazione, scopo dello spostamento e itinerario seguito, mentre le caratteristiche socio-economiche di ciascun partecipante le ricavano dall'analisi delle risposte ai questionari. Lo studio analizza un totale di 601 spostamenti, di cui 212 in auto e 195 in moto. Definendo come percorso principale quello maggiormente utilizzato durante il periodo di rilievo, trovano che circa il 20% degli automobilisti e il 6% di motociclisti, tra la stessa origine e destinazione, utilizzano solo un percorso, mentre i restanti ne utilizzano almeno due; nell'intervallo orario di punta del mattino, AM, solo il 34,5% degli automobilisti usa solo un percorso, contro il 25,9% del PM.

Nella maggioranza dei casi illustrati nelle righe precedenti, i dati GPS sono stati raccolti attraverso delle indagini sui veicoli, che permettono di raccogliere informazioni oggettive sugli spostamenti (tempi e distanze), mentre per ciò che riguarda l'acquisizione delle informazioni legate alle caratteristiche soggettive degli utenti sono effettuate interviste pre e post analisi. Per lo studio dei percorsi vengono utilizzate delle piattaforme GIS. Il limite delle indagini di questo tipo è legato alle poche informazioni acquisibili attraverso il GPS, i cui dati necessitano poi di essere elaborati per estrarre altre caratteristiche, quali gli istanti di inizio, fine. Inoltre, gli spostamenti effettuati non possono essere direttamente legati alle attività compiute dagli utenti, fattore che diminuisce il numero di informazioni aggiuntive a disposizione per la ricerca. Ancora, l'utilizzo di piattaforme GIS per il calcolo dei percorsi minimi non permette di tenere conto adeguatamente delle caratteristiche della mobilità e delle condizioni reali del traffico (semafori, congestione, etc.).

Nella tabella posta di seguito viene riportata una sintesi che confronta le principali caratteristiche delle diverse ricerche analizzate.

	<b>Jan 2000</b>	<b>Li 2005</b>	<b>Parkany 2006</b>	<b>Papinski 2009</b>
<b>Drivers</b>	216	182	106	31
<b>Households</b>	100	138	276	
<b>Trips in the sample</b>	+ 3000		4494	21
<b>Data collection period</b>	1 week	10 days	10 days	2 days
<b>Simulation tool</b>	GIS	GIS	/	GIS
<b>Simulated path parameters</b>	Shortest time and its distance	/	Shortest time	time, distance
<b>Real path parameters</b>	time, distance	time, distance, speed, stops, roads, signals	time, distance, signals, turns, roads	time, distance
<b>choice model</b>	no	si	si	no
<b>GPS device</b>	in vehicle	in vehicle	in vehicle	mobile
<b>Path's time period</b>	daily	5:00 - 10:00	daily	/
<b>% same path - same OD</b>	/	40%	/	/
<b>% + paths - same OD</b>	/	60%	/	/
<b>% routine paths</b>	/	33%	/	/
<b>Area</b>	Lexington, Kentucky	Atlanta	Lexington, kentucky	Ontario, Canada
<b>Diverters</b>	/	/	33 users	/
<b>Trip purpose</b>	/	home/work	/	home/work
<b>Average distance</b>	/	15.9 mph	/	13.9 km
<b>Average time</b>	/	37.4 min	/	17.5 min
<b>Parameters calculated</b>	PDI	/	TDI*	/
<b>TDI values</b>	/	/	> 10% 33 users	/
<b>DDI values</b>	/	/	/	/
<b>OI values</b>	/	/	/	/
<b>RDI values</b>	/	/	/	/

*\* differenti formulazioni*

Tabella 1 - Studi basati su dati GPS

	<b>Zhu 2009</b>	<b>Zhu 2010</b>	<b>Papinsky 2010</b>	<b>Spissu 2011</b>	<b>Arifin 2011</b>
<b>Drivers</b>	35	95		12	93
<b>Households</b>					
<b>Trips in the sample</b>	351	657	237	393	601
<b>Data collection period</b>	8 weeks	13 weeks	48 h	2 weeks	1 week
<b>Simulation tool</b>	GIS	GIS	RCA toolkit/GIS	CUBE/GIS	/
<b>Simulated path parameters</b>	Shortest: time or distance	time, distance, speed	everything	time, distance	/
<b>Real path parameters</b>	speed, distance	time, distance, speed	everything	time, distance	time, distance
<b>choice model</b>	no	si	no	no	no
<b>GPS device</b>	in vehicle	in vehicle	mobile	mobile	in vehicle
<b>Path's time period</b>	morning peak	morning peak	daily	daily	daily
<b>% same path - same OD</b>	32.5%	60%	/	92.7%	26.0%
<b>% + paths - same OD</b>	67.5%	40%	/	7.3%	74.0%
<b>% routine paths</b>	/	/	/	/	/
<b>Area</b>	Twin Cities, Minnesota	Twin Cities, Minnesota	Halifax, Canada	Cagliari, Italy	Jakarta
<b>Diverters</b>	/	/	/	/	/
<b>Trip purpose</b>	home/wor k	home/work	home/work	all	/
<b>Average distance</b>	/	/	/	4.87 km	/
<b>Average time</b>	/	/	/	11.76 min	/
<b>Parameters calculated</b>	TDI*, DDI*	/	RDI	TDI*, DDI*, OI	/
<b>TDI values</b>	> 50%	/	/	> DDI - < 4 km	/
<b>DDI values</b>	> 50% per il 5%	/	/	>TDI - > 7 km	/
<b>OI values</b>	/	/	/	30% > 0.6	/
<b>RDI values</b>	/	/	1.53	/	/

*\* differenti formulazioni*

Tabella 1 - Studi basati su dati GPS

### 2.1.1. Intravariabilità

Attraverso il monitoraggio degli spostamenti con i dispositivi GPS, è stato possibile rilevare anche alcuni fenomeni particolari legati al comportamento degli utenti, tra cui l'intravariabilità degli spostamenti,

Come anticipato nell'introduzione, questo particolare aspetto, noto in forma generale come route switching, è stato analizzato secondo due modi: nella prima, quella relativa all'analisi dei dati GPS, gli autori hanno semplicemente messo in evidenza il particolare comportamento, senza però analizzarlo in dettaglio; nella seconda gli autori provano a stimare alcuni modelli di scelta discreta in modo da individuare quali siano le variabili che maggiormente influenzano tale fenomeno. In quest'ultimo caso, tuttavia, i dati utilizzati derivano da indagini su stated preferences (SP) e sono quasi tutte volte ad indagare sugli effetti derivanti dalla somministrazione di informazioni di percorso. Un breve riepilogo di questa seconda categoria di studi è riportata nel seguito, in quanto costituisce la base di partenza delle svolte, che si pongono come obiettivo quello di "fondere" i due filoni (GPS e studio modellistico).

Mannering (1989) ha usato una regressione di Poisson per predire la frequenza mensile di cambi di percorso degli utenti pendolari, trovando come gli attributi specifici della rete risultati più significativi siano la disponibilità di percorsi alternativi, il livello di congestione e il tempo di viaggio del percorso principale, mentre tra le variabili socioeconomiche un ruolo importante lo occupano l'età e lo stato civile.

Mahmassani, et al. (1990) sviluppa un modello logit binario che lega il cambio di percorso alle condizioni geografiche e della rete, caratteristiche del posto di lavoro, attributi individuali e uso delle informazioni (rapporti radio sul traffico), trovando molto significative le variabili proprie dello spostamento stesso, unitamente all'uso delle informazioni.

Khattak, et al. (1991) analizzano l'effetto dei rapporti sul traffico in relazione al cambio di percorso e all'istante di partenza, utilizzando dati derivati da un questionario di tipo "mail-back". Dall'applicazione di un modello ordered probit, trovano come variabili significative, oltre a quelle direttamente legate ai rapporti sul traffico, la congestione, la differenza tra i tempi di viaggio, il numero di volte che è stato scelto il percorso abituale e il numero di percorsi alternativi a disposizione.

Mannering & Kim (1994) applicano un modello ordered logit a dati collezionati attraverso un questionario nell'area metropolitana di Seattle, per predire la frequenza di cambio percorso nei tragitti casa-lavoro: trovano che quest'ultima aumenti all'aumentare della durata del tragitto. Mostrano inoltre che gli utenti aventi dichiarato buona familiarità col percorso e flessibilità negli orari di entrata e uscita da lavoro siano più frequenti "route changers",

mentre le donne e gli individui con un basso reddito siano invece meno predisposti al cambio.

Polydoropoulou, et al. (1996) si pongono l'obiettivo di analizzare le decisioni prese durante lo spostamento nel momento in cui, agli utenti, sono fornite delle informazioni per mezzo di sistemi ATIS, esplorando le reazioni degli stessi in caso di congestione inaspettata e in cui l'informazione sia di tipo qualitativo, quantitativo, previsiva o prescrittiva. Utilizzano dati raccolti attraverso un questionario in cui si chiedevano informazioni sul percorso realmente fatto (RP) e dove inoltre si proponevano scenari alternativi (SP). Stimano un modello basato di scelta discreta, trovando che le variabili che hanno maggiore importanza nel cambio di decisione siano il ritardo atteso nel percorso abituale, il tempo di viaggio dei percorsi alternativi, il livello di congestione percepito e le sorgenti di informazione (perché nel momento in cui osservano il ritardo è troppo tardi per cambiare).

Lam (2000) sviluppa un modello teorico per analizzare le decisioni riguardanti l'istante di partenza e la scelta del percorso in una rete semplice caratterizzata da due itinerari paralleli: uno gratuito ma congestionato e l'altro scarico ma a pagamento. Lo strumento utilizzato consente di valutare due tipologie di incertezze: la lunghezza del tragitto e la tipologia di percorso. Nello studio viene osservato come per spostamenti di lunga distanza gli utenti siano meno propensi a usare il tratto a pagamento, mentre quelli che pianificano in anticipo di usare le toll lanes mostrano meno disponibilità al cambio perché l'istante di partenza è già stato influenzato dalle considerazioni sul tempo di viaggio risparmiato e la minore incertezza dello stesso.

Han, et al. (2001) utilizzano dati ricavati da un'indagine SP per testare diverse specificazioni modellistiche, con lo scopo di individuare quella migliore per l'analisi del cambiamento di percorso, mostrando come i mixed logit siano quelli con la struttura più adatta. Trovano che le variabili più significative siano i ritardi, la non-conoscenza del ritardo e la poca disponibilità a fidarsi del valore dello stesso, il tempo di percorrenza a rete congestionata, il sesso, l'essere dei frequenti guidatori e la volontà di evitare la congestione.

Srinivasan & Mahmassani (2003) indagano sulle dinamiche del cambio di percorso nel caso di disponibilità di sistemi informativi avanzati (ATIS), applicando dei modelli di tipo logit kernel (mixed logit) dinamici a dati ricavati attraverso un esperimento in laboratorio, con la possibilità di tenere in considerazione l'eterogeneità delle scelte e la correlazione tra le stesse. Oltre alle variabili direttamente legate alla tipologia di informazione somministrata, particolare significatività assumono il risparmio di tempo, il livello di congestione, il costo del cambio (in miglia).

Jou (2004) utilizza un modello che combina equazioni di misura e strutturali, con variabili latenti, per studiare l'impatto che le informazioni in tempo reale sul traffico hanno sul

cambio di percorso, analizzando dati raccolti attraverso la somministrazione di un questionario. Stima poi dei modelli logit binari, con lo stesso scopo, trovando come significative età, abitudine, lunghezza, frequenza e urgenza dello spostamento.

Xu, et al., (2010) Raccoglie i dati sulla scelta del percorso attraverso un'indagine SP nella quale si chiedeva agli utenti di scegliere, sotto differenti condizioni di viaggio (istante di partenza, tempi di viaggio, variazioni del tempo e congestione), tra tre differenti itinerari, all'inizio dei quali venivano fornite delle informazioni di percorso, utilizzando un MNL per le analisi di scelta discreta. Trova che le variabili più significative sono l'età, gli anni di esperienza alla guida, il tempo di viaggio e la sua variazione.

Tawfik, et al. (2010) effettuano delle sperimentazioni attraverso un simulatore di guida e un questionario pre e post rilevazione, analizzando statisticamente l'evoluzione delle scelte di percorso basandosi sulle caratteristiche degli utenti monitorati, con lo scopo di valutare le differenze di comportamento in relazione alla variazione delle performance della rete e al grado di apprendimento degli individui.

Ben-Elia & Shiftan (2010) conducono un esperimento in laboratorio per l'applicazione di un modello basato sull'apprendimento (learning based model), con l'obiettivo di interpretare il meccanismo di scelta in caso di informazioni (in particolare sul tempo di viaggio) fornite in tempo reale. Usano una specificazione di tipo mixed logit con panel data.

Albert, et al. (2011) effettuano un esperimento in laboratorio per valutare l'impatto che diversi aspetti della personalità individuale hanno nel comportamento di scelta del percorso, in caso di presenza di informazioni sul tempo di viaggio parziali fornite prima di effettuare lo spostamento. Trovano che gli individui con scarse abilità spaziali tendano a cambiare di meno, mentre quelli che hanno ottenuto un punteggio alto nella ricerca di buone sensazioni (bisogno di sperimentare situazioni ed esperienze nuove, varie e complesse) mostrano una buona frequenza di cambio percorso. Per le stime di scelta discreta usano un modello logit di regressione, trovando buona significatività per le variabili legate al tempo di viaggio, abilità spaziale, intolleranza verso routine di qualsiasi tipo, frequenza di spostamento, sesso.

<b>Autori</b>	<b>Anno</b>	<b>Indagine</b>	<b>Metodo</b>	<b>Variabili significative</b>
Mannering	1989	Questionario RP	Regressione di Poisson	Disponibilità di percorsi alternativi, livello di congestione, tempo di viaggio nel percorso principale, età, stato civile
Mahmassani et al.	1990	Questionario RP	Logit binario	Variabili proprie dello spostamento (tempi, distanze), utilizzo delle informazioni
Khattak et al.	1991	Questionario RP	Probit ordinato	Variabili legate ai rapporti sul traffico, congestione, differenza dei tempi di viaggio, numero di volte che il percorso abituale è stato scelto, numero di percorsi alternativi
Mannering & Kim	1994	Questionario RP	Logit ordinato	Durata del tragitto, familiarità col percorso, flessibilità negli orari di ingresso e uscita da lavoro, sesso, reddito
Polydoropoulou et al.	1996	Questionario RP e SP	Modello basato sulla RUM	Ritardo atteso nel percorso principale, tempo di viaggio, livello di congestione percepito, sorgenti di informazione
Han et al.	2001	Questionario SP	Logit misto	Ritardi, non-conoscenza del ritardo, poca propensione a ritenere affidabile il valore di ritardo, sesso, essere o meno un guidatore frequente, volontà di evitare la congestione
Srinivasan & Mahmassani	2003	Laboratorio	Logit kernel	Tipologia di informazione, risparmio di tempo, livello di congestione, "costo" (in miglia) del cambio di percorso
Jou	2004	Questionario RP	Logit binario	Età, abitudine, lunghezza, frequenza e urgenza dello spostamento
Xu et al.	2010	Questionario SP	Logit multinomiale	Età, anni di esperienza alla guida, tempo di

				viaggio e la sua variazione
Ben-Elia & Shiftan	2010	Laboratorio	Logit misto - panel	Variabili legate al tempo di viaggio
Albert et al.	2011	Laboratorio	Regressione logit	Tempo di viaggio, abilità spaziale, intolleranza verso routine di qualsiasi tipo, frequenza dello spostamento, sesso

Tabella 2 - Riassunto degli studi sul cambio di percorso

Le risultanze derivate da studi di questo tipo hanno una validità limitata, in quanto le preferenze osservate possono essere influenzate dalla mancanza di realismo e dalla percezione soggettiva, per ciascun utente, dell'astrattezza della situazione di scelta (Carrion & Levinson, 2012). Sono stati individuati solamente due studi nei quali tali argomenti sono stati affrontati utilizzando una base dati costruita a partire da un'indagine basata sul sistema GPS. Zhang & Levinson (2008) utilizzano dati rilevati attraverso un dispositivo GPS installato all'interno del veicolo. Il campione è costituito da 113 guidatori a cui si è chiesto di effettuare degli spostamenti su quattro percorsi alternativi tra delle preselezionate coppie origine/destinazione, e a cui vengono fornite, in tempo reale, delle informazioni di percorso con diverso grado di accuratezza. Analizzano i dati attraverso l'applicazione di modelli come la regressione ordinaria, logit multinomiale e logit a rango ordinato (rank-ordered), stimando l'effetto dovuto alla somministrazione delle informazioni sui percorsi (Zhang & Levinson, 2008). Li, et al. (2005), invece, utilizzano i dati GPS, relativi a spostamenti sistematici avvenuti nell'ora di punta del mattino, proprio per valutare quali siano gli attributi del percorso e le caratteristiche socioeconomiche degli utenti che rivestono maggiore importanza nel determinare che un utente sia un "changer" oppure utilizzi sempre lo stesso percorso. Stimano quattro modelli logit binari, di cui uno che include solo le informazioni dello spostamento, il secondo le caratteristiche del percorso principale (più utilizzato durante il rilievo), il terzo le informazioni socio-demografiche e il quarto tutte le variabili contemporaneamente. Dai risultati si evince che la maggiore significatività sia da ricercare nel numero di stop intermedi durante lo spostamento, nella flessibilità dell'orario d'arrivo a lavoro e nell'età dell'utente.



## **2.2. La scelta del percorso**

### **2.2.1. Panoramica generale sul processo decisionale**

La rappresentazione del processo decisionale ha mosso i primi passi agli inizi del 19° secolo, quando gli economisti neoclassici iniziarono a sviluppare i primi modelli matematici, i quali erano basati sull'ipotesi che gli individui siano razionali, perfettamente informati sulle alternative a loro disposizione e interessati autonomamente ad incrementare il proprio benessere, secondo la teoria meglio conosciuta come dell'uomo economico. Tali ipotesi semplificative sono necessarie nella modellazione dell'universo dei fenomeni sociali, perché altrimenti, a causa della vastità e complessità delle componenti che li contraddistinguono, i modelli di rappresentazione rischiano di diventare molto rapidamente intrattabili dal punto di vista matematico. Con il passare del tempo e con l'aumentare delle ricerche e sperimentazioni, è nata la necessità di poter applicare tali modelli all'analisi di contesti di scelta sempre più ampi e diversamente caratterizzati, rendendo necessario un continuo processo evolutivo degli strumenti di rappresentazione, fino ad arrivare alla specificazione del modello ad utilità additiva, sviluppato a partire da quello dell'uomo economico con lo scopo di descrivere le scelte tra alternative secondo schemi multi-attributo. L'ipotesi fondamentale alla base dell'intero approccio teorico è quella che suppone che l'individuo sia perfettamente informato sulle alternative a disposizione, conosca esattamente tutte le caratteristiche e particolarità di ciascuna di esse e effettui le proprie scelte in modo da massimizzare la propria utilità, valutata come somma pesata dei valori associati ad ogni singolo attributo dell'alternativa in accordo con la struttura di preferenza individuale. Una caratteristica importante del modello ad utilità additiva è la sua natura puramente compensativa, in quanto le alternative possono essere assimilate a dei contenitori di utilità, nei quali un alto valore di un attributo può compensare quello basso di un altro. Le differenze riscontrate tra il comportamento reale rispetto a quello modellato secondo i principi dell'utilità additiva furono considerati dei puri errori di valutazione, di cui si tenne conto nello sviluppo della teoria dell'utilità probabilistica (Luce, 1959; Block & Marschak, 1960). Simon (1955) suggerì invece che le discrepanze tra il comportamento reale e quello rappresentato secondo la struttura modellistica dell'utilità additiva derivassero da vincoli esterni imposti dal contesto di scelta, come il tempo e la disponibilità di informazioni, unitamente ai limiti tipici dell'individuo, come l'abilità ad elaborare le informazioni, capacità computazionali e previsionali, che portano ad avere delle percezioni distorte riguardo agli eventi o al comportamento delle altre persone, o a distorsioni nella scelta che si traducono in differenze rispetto a quella che teoricamente soddisferebbe le proprie preferenze. A partire da queste considerazioni, Simon sviluppò la teoria della razionalità vincolata (bounded rationality), secondo la quale l'utente applica delle strategie decisionali di semplificazione

(euristiche) del panorama di scelta sulla base delle proprie capacità e limiti individuali. In pratica, le euristiche consentono alla persona di costruire rapidamente una propria rappresentazione della realtà, all'interno della quale effettueranno ragionamenti più avanzati di scelta.

In generale, in tutti gli approcci che hanno l'obiettivo di rappresentare il comportamento individuale che sta alla base di tutti i modelli di previsione, una scelta può essere vista come il risultato di un processo decisionale sequenziale definito come segue:

1. Definizione del problema di scelta;
2. Generazione delle alternative;
3. Valutazione degli attributi delle alternative;
4. Confronto tra le alternative sulla base delle loro caratteristiche;
5. Scelta;
6. Implementazione.

Nello specifico, una teoria di scelta è costituita da un insieme di procedure che definiscono i seguenti elementi:

- Chi deve decidere;
- Alternative;
- Attributi delle alternative;
- Regole decisionali.

È importante sottolineare come non tutte le scelte osservate siano frutto di un processo decisionale esplicito. Ad esempio, un individuo può seguire un'abitudine, assumere delle forme di comportamento convenzionali, affidarsi all'intuito oppure imitare qualcuno considerato esperto o leader. L'ambiente nel quale l'individuo opera determina quello che è generalmente conosciuto come insieme universale delle alternative, del quale però sarà possibile considerarne solo un sottoinsieme chiamato insieme di scelta. Infatti, la disponibilità e utilizzabilità di un'alternativa è definita da una varietà di limitazioni e vincoli, come la disponibilità fisica (ad esempio l'esistenza di un servizio di autobus nel percorso casa-lavoro), risorse monetarie (possibilità di prendere un taxi), disponibilità di tempo, limiti legati all'informazione, esperienza e tutti gli schemi mentali che contraddistinguono i singoli individui e fanno sì che essi possano prenderla in considerazione.

Per effettuare una scelta da uno di questi insiemi, nel momento in cui sono formati da due o più alternative, è richiesta una regola decisionale che descriva il meccanismo interno che il decisionista attiva per processare le informazioni disponibili e arrivare ad una scelta univoca. In letteratura ne esistono una varietà, classificabili nelle seguenti 4 categorie:

- Dominanza: un'alternativa è considerata dominante rispetto ad un'altra se è migliore per almeno un attributo e non inferiore rispetto agli altri, nel rispetto di alcune regole di classificazione (ad esempio, insieme di valori per i quali la classificazione non cambia);
- Soddisfazione: tutti gli attributi devono avere un valore compreso entro certe soglie;
- Regole lessicografiche: gli attributi sono classificati in ordine di importanza e si sceglie l'alternativa più attrattiva secondo l'attributo più importante. La combinazione tra le regole soddisfazione e lessicografica è quella denominata eliminazione per aspetti;
- Utilità: l'attrattività degli attributi diventa uno scalare e la scelta avviene secondo una regola di massimizzazione.

Una delle prime regole decisionali ad essere stata sviluppata è quella di soddisfazione (satisfaction), suggerita da Simon (1955), con il fine di effettuare un'approssimazione del processo di semplificazione del panorama di scelta di fronte a condizioni di incertezza: seguendo uno schema di ricerca sequenziale, un individuo sarebbe disposto ad optare per un'alternativa le cui caratteristiche superano le soglie di accettabilità da lui fissate, invece di scegliere quella che, razionalmente, sarebbe stata quella migliore. Infatti, spesso le persone non si formano delle convinzioni razionali perché, sistematicamente, dei dettagli apparentemente irrilevanti influiscono sul loro comportamento e non scelgono sempre quello che sembra essere il loro interesse principale (Camerer & Fehr, 2006).

Un'altra delle prime regole studiate è quella della differenza additiva (Tversky, 1969), nella quale le alternative vengono confrontate a coppie, sulla base delle caratteristiche degli attributi di ciascuna, i cui valori vengono valutati e sommati al fine di effettuare una valutazione globale. Di seguito si riporta un breve elenco di regole presenti in letteratura (Hauser, et al., 2009; Tversky, 1969; Einhorn, 1970; Tversky, 1972; Tversky, 1972; Payne, et al., 1993):

- Lessicografica – il consumatore prima classifica le caratteristiche delle alternative, poi ordina queste sulla base della caratteristica a cui è associato il valore più elevato, in un processo a cascata; l'ordinamento avviene su scala decrescente e la scelta si effettua in base alla posizione in classifica;
- Congiuntiva – un'alternativa è presa in considerazione solamente se tutti gli attributi, per i quali il decisore stabilisce i livelli minimi di accettabilità, hanno dei valori che rientrano in determinate soglie;
- Disgiuntiva – un'alternativa è presa in considerazione se il valore di almeno una caratteristica supera una determinata soglia;

- Sottoinsieme congiuntivo – un’alternativa deve avere almeno  $S$  caratteristiche che soddisfano determinati criteri. È la generalizzazione della regola congiuntiva ( $S$ =tutte le caratteristiche) e disgiuntiva ( $S=1$ );
- Disgiunzioni di congiunzioni (DOC) – un’alternativa è considerata se una o più congiunzioni sono soddisfatte. È la generalizzazione ulteriore delle tre precedenti;
- Eliminazione per aspetti (EBA) – si selezionano in sequenza le caratteristiche indesiderate e, di volta in volta, si elimina l’alternativa che le contiene;
- Maggioranza delle caratteristiche confermate – in un confronto a coppie viene scelta l’alternativa che ha il maggior numero di attributi migliori, senza tenere conto del peso degli stessi o delle differenze;
- Frequenza di caratteristiche buone o cattive – si basa sul conteggio delle caratteristiche positive e negative.

### **2.2.2. Il processo decisionale nella scelta del percorso**

Anche la scelta del percorso, dato il ruolo principale che la componente comportamentale assume all’interno dell’intero processo mentale dell’individuo, si inserisce nel campo della teoria delle decisioni. Data la notevole complessità associata alla sua rappresentazione e alla ridotta disponibilità di dati reali sui quali effettuare le sperimentazioni, per via delle difficoltà relative alla raccolta degli stessi, la scelta del percorso è soggetta a numerose semplificazioni sia dal punto di vista teorico che applicativo. Infatti, il problema dell’assegnazione del traffico è stato maggiormente analizzato dal punto di vista del caricamento delle reti e nella soluzione degli schemi di flussi sugli archi nelle aree urbane congestionate, riservando un’attenzione marginale alla componente comportamentale alla base della scelta del percorso (Prashker & Bekhor, 2004). Una corretta modellazione della stessa, invece, darebbe la possibilità di valutare le percezioni che guidatori hanno sulle caratteristiche del percorso, di prevederne il comportamento in particolari situazioni, di effettuare previsioni sulle future condizioni del traffico e di prevedere le reazioni che gli individui avrebbero di fronte a sistemi di informazione (Prato, 2009).

Nelle applicazioni pratiche nel campo dell’ingegneria e della pianificazione, come ad esempio l’assegnazione dinamica del traffico, i sistemi di controllo avanzati di area vasta, i sistemi di informazione ai viaggiatori e i sistemi di guida assistita, i modelli di scelta del percorso costituiscono una componente essenziale dei modelli di assegnazione del traffico (Peeta & Ziliaskopoulos, 2001; Rakha & Tawfik, 2009). È ben documentato, nella letteratura riguardante le scienze psicologiche, che gli individui tendano a minimizzare i propri sforzi cognitivi e seguano semplici euristiche per prendere le proprie decisioni, specie in condizioni di incertezza e vincoli dove le abilità cognitive diventano abituali e tendono a diventare

automatiche, minimizzando quindi le risorse cognitive (Wickens & Hollands, 1999; Tawfik, et al., 2010). Nei modelli che seguono la teoria dell'equilibrio deterministico dell'utente (DUE), il modello di scelta del percorso è costruito sull'ipotesi che gli utenti abbiano una perfetta conoscenza della rete, dei costi di percorso e si comportino identicamente tra loro, scegliendo quello che minimizza il costo di viaggio (Prashker & Bekhor, 2004; Prato, 2009). In questi modelli, la congestione è rappresentata attraverso vincoli di capacità e l'equilibrio dell'utente è trovato in accordo col primo principio di Wardrop (1952), il quale indica che tutti i guidatori scelgano il percorso di minimo costo e si raggiunge l'equilibrio quando nessun utente può ottenere un risparmio cambiando percorso. Nonostante i modelli DUE siano quelli più usati nelle applicazioni pratiche, sono caratterizzati da alcuni limiti: infatti, studi empirici hanno dimostrato che i vincoli di capacità sono insufficienti per spiegare la varietà dei percorsi scelti (Maher & Hughes, 1998). Quindi, le assunzioni alla base del DUE sono irrealistiche, perché in realtà la scelta è basata sui costi percepiti che sono variabili tra gli individui e spesso con ragioni che non sono catturate dalle funzioni di costo (Gliebe, et al., 1999). Partendo quindi dall'inadeguatezza della rappresentazione del comportamento, sono stati sviluppati dei modelli stocastici (Burrell, 1968; Dial, 1971) per tenere conto della variabilità delle percezioni dei tempi e dei costi, utilizzando una distribuzione probabilistica per identificare i costi sugli archi. Poiché questi metodi non permettono di considerare i vincoli di capacità, limitandone quindi l'applicabilità alle reti non congestionate, furono sviluppate delle teorie che combinassero le tecniche DUE con l'approccio stocastico, generando così i modelli SUE (equilibrio stocastico dell'utente), i cui primi sviluppatori furono (Daganzo & Sheffi, 1977; Fisk, 1980; Sheffi, 1985); la scelta del percorso è modellata come un processo stocastico, mentre i vincoli di capacità sono rappresentati attraverso funzioni di costo relative agli archi. Le prime specificazioni utilizzate furono quella MNL (Multinomial Logit) e MNP (Multinomial Probit). (Fisk, 1980) sviluppò una formulazione usando una funzione di entropia, specifica per l'applicazione nella struttura MNL, la cui convenienza analitica ha ispirato diversi altri algoritmi, ad esempio quello di Dial (2001). Un altro ramo della ricerca ha riguardato lo sviluppo dei modelli di assegnazione dinamica del traffico (DTA), i cui primi studi risalgono a partire dalla seconda metà degli anni 50, (i modelli *LWR* (Lighthill & Whitham, 1955; Richards, 1956)). In seguito, molti autori hanno effettuato degli studi che rappresentano delle estensioni dei modelli macroscopici spazio continui *LWR* (si citano ad esempio (Payne, 1971; Papageorgiou, et al., 1990; Daganzo, 1994)). Merchant & Nemhauser (1978) sono stati i primi ricercatori a trattare il problema secondo un approccio macroscopico spazio discreto, e diversi autori hanno fatto riferimento ai loro studi per proporre nuove formulazioni (si citano ad esempio Carey (1992) e Friesz et al. (1993)). Solo in tempi più recenti, invece, sono state sviluppate ricerche sui modelli dinamici con un approccio microscopico (Van Aerde, 1994; Ben-Akiva, et al., 1997), e

mesoscopico (Leonard, et al., 1989; Mahmassani, et al., 1992; Ben-Akiva, et al., 1998; Mahut, et al., 2003), per citarne alcuni). In questi contesti, il modello di scelta del percorso può essere predeterminato oppure calcolato dinamicamente in pari passo con il caricamento della rete (Prato, 2009).

Dalla breve panoramica effettuata riguardo gli strumenti operativi di simulazione, è emerso che la scelta del percorso è rappresentata secondo schemi semplificati e non completamente esaustivi in merito alla riproduzione della realtà. Infatti, in tutti gli approcci analizzati, le caratteristiche che vengono considerate negli algoritmi di calcolo tengono conto degli attributi oggettivi dei percorsi e del traffico sulla rete, come la capacità, la distanza, il tempo di percorrenza. Negli approcci stocastici si evidenzia uno sforzo maggiore nella rappresentazione della componente comportamentale, ma anche in questo caso ci si limita all'implementazione di funzioni probabilistiche applicate ai costi sugli archi, che non sono quindi legate ad aspetti più strettamente individuali correlati alle preferenze di scelta.

Queste semplificazioni, tuttavia, sono rese necessarie dalle caratteristiche dell'ambiente di scelta con il quale ci si interfaccia nel caso della scelta del percorso. Diversi studi dedicati al processo decisionale hanno mostrato che, in situazioni nelle quali l'individuo si trova di fronte a molte alternative, si attiva un processo decisionale costituito da una sequenza a due stadi: la prima fase è puramente non compensativa ed è costituita dall'applicazione di regole euristiche basate sugli attributi che permettono di selezionare le opzioni disponibili; la seconda fase, costituita da una strategia compensativa, si basa sugli attributi dell'alternativa (Payne, 1976; Olshavsky, 1979; Payne, et al., 1993) e rappresenta la decisione finale. La scelta del percorso, proprio per la natura dell'ambiente di scelta costituito dalle reti urbane di trasporto, fa parte della casistica di situazioni nelle quali l'individuo si interfaccia con un elevato numero di alternative, ragione per cui è considerato un processo a due stadi costituito dalla formazione del set di scelta (a partire dall'universo generale delle opzioni a disposizione) e quindi dalla scelta finale.

Il modo di agire nella scelta del percorso è quindi un argomento complesso e molto sfaccettato, in quanto la sua interpretazione cambia se si considera il punto di vista del viaggiatore oppure quello del ricercatore. Nei paragrafi successivi, usando come rappresentazione base il punto di vista dell'individuo così come evidenziato da Bovy & Stern (1990) e Bovy (2009), si analizza poi il punto di vista del ricercatore, il quale deve elaborare teorie e applicazioni pratiche funzionali alla rappresentazione dell'intero processo decisionale.

### 2.2.3. Punto di vista dell'individuo

L'individuo, a causa delle limitazioni dovute alla conoscenza della rete, alle abitudini consolidate e agli schemi mentali, nel momento in cui deve selezionare il percorso da utilizzare non considera tutti quelli esistenti. Generalmente, infatti, le percezioni dell'individuo sulle alternative rilevanti e i loro attributi risultano essere incomplete, inaccurate e legate alle proprie esperienze e preferenze. Ad esempio, alcuni percorsi non sono percepiti come distinti a causa della mutua sovrapposizione. Le alternative conosciute da ciascun individuo andranno a formare l'insieme di scelta da cui poi attingerà per prendere la decisione finale, che non dipenderà necessariamente dalle caratteristiche direttamente misurabili del percorso ma anche da fattori di scelta personali come costi, comfort, numero di semafori, numero di svolte, preferenze di viaggio. La figura seguente mostra la struttura del processo decisionale della scelta del percorso vista dalla prospettiva dell'individuo (Bovy & Stern, 1990), dove le frecce mostrano come tutto il processo di formazione dell'insieme scelta sia un meccanismo di apprendimento in evoluzione che avviene attraverso adattamenti dinamici delle capacità cognitive e percettive sulle alternative disponibili e valutabili.

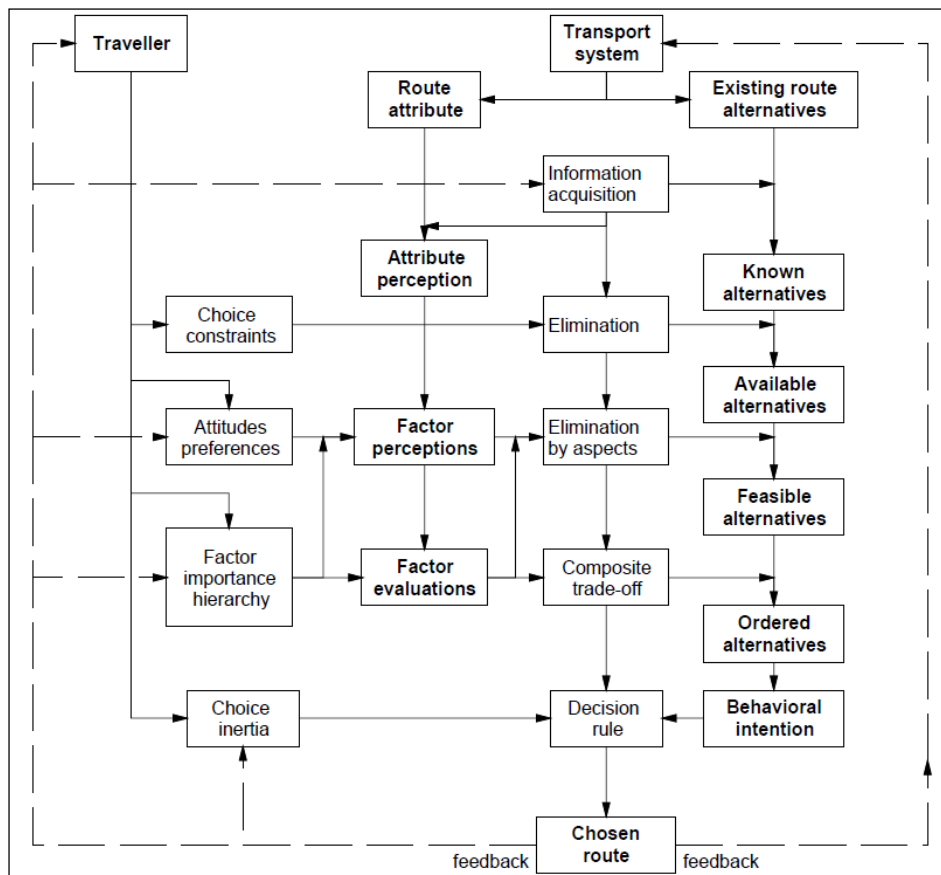


Figura 1 - Processo decisionale nella scelta del percorso (Bovy & Stern, 1990)

Secondo questo processo, l'individuo considera le alternative a sua disposizione secondo vincoli di scelta, attitudini, preferenze e fattori di importanza gerarchica. Dal sistema di trasporti si estraggono le opzioni a disposizione riguardanti i percorsi esistenti, che l'utente non conosce del tutto e, grazie alle informazioni precedentemente acquisite, seleziona il proprio insieme di alternative conosciute. Le informazioni a sua disposizione e i vincoli di scelta gli consentono di eliminare, da questo insieme, le opzioni che non lo soddisfano, arrivando alla formazione dell'insieme dei percorsi effettivamente disponibili. La percezione degli attributi di percorso, legata alle attitudini e alle preferenze individuali, consente di attivare il processo di eliminazione per aspetti (euristica), attraverso il quale sarà possibile la formazione dell'insieme delle alternative effettivamente utilizzabili. Attraverso un meccanismo di valutazione degli attributi, in base a quanto già sperimentato nel passato e alle inclinazioni di scelta, l'individuo applicherà poi una regola decisionale con la quale selezionerà l'alternativa finale.

Riassumendo, nelle reti di trasporto l'insieme universale delle alternative non è conosciuto nella sua interezza, sia per l'elevata quantità di percorsi sia per le limitate capacità cognitive dell'individuo, il cui insieme disponibile sarà determinato dalle proprie preferenze di viaggio. Egli prenderà poi in considerazione, da questo, solamente quei percorsi che in quel momento soddisferanno le proprie esigenze, ottenendo l'insieme utilizzabile. La scelta finale avverrà effettuando delle valutazioni più dettagliate attraverso l'applicazione di svariate regole decisionali, tradotte con la collocazione di tutti i fattori di scelta in una scala comune, chiamata utilità, che dipenderà poi dai valori soggettivi che egli attribuirà alle caratteristiche, osservabili e non, del percorso.

#### **2.2.4. Punto di vista del ricercatore**

Nel paragrafo precedente si è parlato di come il viaggiatore ragioni e si comporti nel momento in cui deve effettuare uno spostamento e quindi una scelta di percorso. Ovviamente, per essere riprodotto, tale schema comportamentale necessita di essere tradotto in modo da implementarlo in modelli matematici di simulazione e, per far questo, occorre visualizzare l'argomento dal punto di vista del ricercatore. Infatti vi sono diversi aspetti che, essendo prettamente legati all'individuo, non possono essere conosciuti, come ad esempio l'insieme di scelta da cui poi l'individuo decide il proprio percorso. Questo fattore, oltre a variare tra gli utenti (i quali, nei loro schemi mentali, ne posseggono una visione deterministica) non può essere quantificato o rilevato con nessun sistema. A causa di ciò, questo insieme di scelta porterebbe a semplificazioni necessarie sui modelli da sviluppare, in



quanto si dovrebbero utilizzare regole probabilistiche o algebra fuzzy per poterli definire. Anche l'insieme universale dei percorsi, al quale appartengono tutti gli itinerari possibili tra una certa origine e destinazione, non può essere conosciuto né dall'analista né dal viaggiatore, a causa della vastità che lo contraddistingue. Nel paragrafo precedente, infatti, si è visto che l'individuo approssima l'insieme universale in uno di alternative disponibili a lui conosciute (chiamato insieme disponibile) e, in seguito, sulla base delle proprie preferenze di viaggio, andrà a comporre l'insieme di percorsi utilizzabili (insieme utilizzabile: quest'ultimo è in genere di ridotte dimensioni e legato alle particolari condizioni dell'individuo e dello specifico viaggio, come la disponibilità del veicolo, limiti di tempo, massima deviazione accettabile e le preferenze).

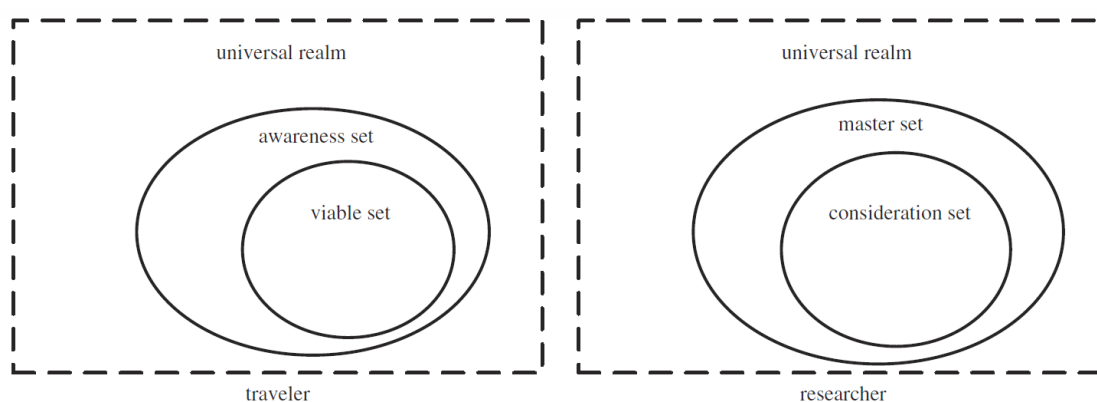


Figura 2 - Prospettiva del viaggiatore e del ricercatore (Kaplan & Prato, 2012)

Quando rappresentano la scelta dei viaggiatori, i ricercatori non possono evidenziare l'insieme universale perché l'enumerazione di tutti i possibili percorsi è praticamente impossibile in reti urbane dense. L'analista ha quindi bisogno di specificare un insieme di percorsi base da cui derivare i possibili utilizzabili, estrapolandoli generalmente da una versione semplificata della rete reale. Gli insiemi utilizzabili estratti in questo modo saranno chiamati, nel seguito, insiemi principali ("master set" per distinguerli da quelli disponibili dell'individuo). Prodotto utilizzando un certo tipo di criteri, questo insieme principale è ben definito dal punto di vista dell'analista e la sua dimensione e composizione può essere conosciuta, nonostante non sia un vero set di scelta in quanto potrebbe incorporare anche alternative non utilizzabili. Solo nel caso di sufficienti informazioni sulle specifiche condizioni del viaggiatore (tipo la sua familiarità con la rete, la frequenza di spostamento tra la stessa coppia OD e le preferenze) il modellista può effettuare delle previsioni probabilistiche dell'insieme utilizzabile, chiamato insieme di considerazione, di cui ha bisogno per la modellazione della scelta del percorso. Queste ipotesi possono andare bene nei casi di stima dei parametri della funzione di costo basate sulle osservazioni individuali del comportamento di viaggio. Dal momento però che le informazioni sulle

specifiche condizioni personali del viaggiatore sono spesso esigue, l'insieme di considerazione solitamente è più grande di quello utilizzabile dell'individuo. Idealmente, l'insieme di considerazione e quello utilizzabile dovrebbero essere perfettamente uguali. Tuttavia, è possibile solamente approssimare quello individuale per via di due principali cause di incertezza: la prima causa è che, anche se i set principali dovessero riprodurre fedelmente l'insieme disponibile dei percorsi conosciuti dai viaggiatori, praticamente la loro somiglianza non è garantita (Bovy, 2009; Prato, et al., 2012); il secondo motivo di incertezza sono le limitazioni spazio-temporali che variano tra la popolazione e possono essere previste solo con un certo grado di accuratezza. Questi aspetti saranno mediati tenendo conto dell'eterogeneità della popolazione e incorporando l'incertezza sia nella generazione del master set che nella formazione dell'insieme di considerazione. La figura seguente riassume quanto detto finora.

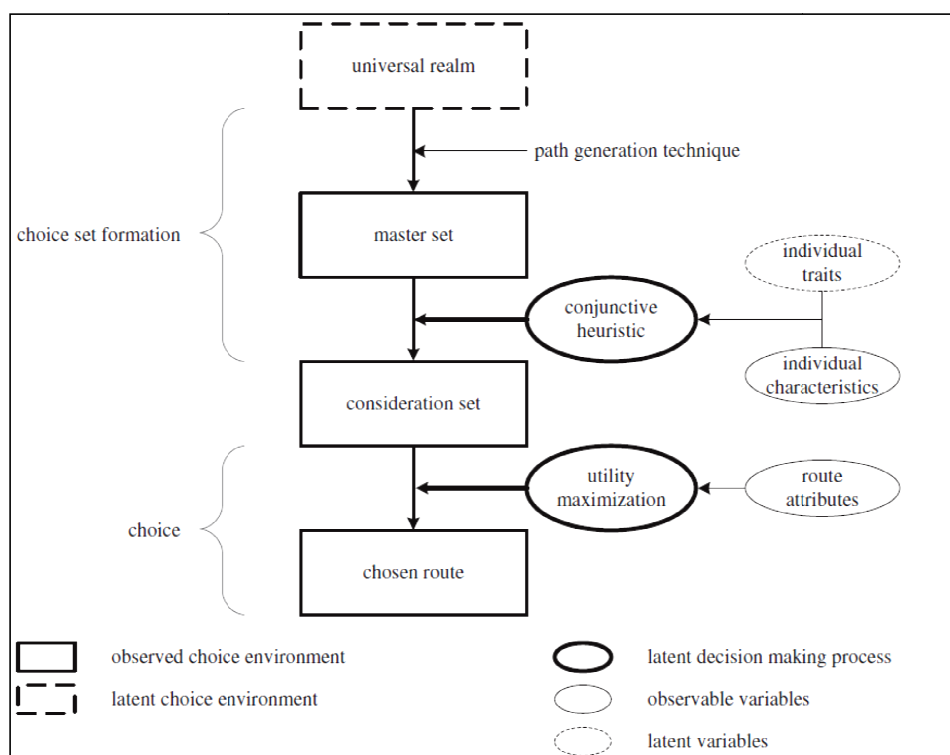


Figura 3 - Modellazione della scelta del percorso (Kaplan & Prato, 2012)

La figura mostra l'intero processo di modellazione della scelta del percorso secondo lo schema a due stadi, composti dalla formazione dell'insieme di scelta e dalla scelta finale. Nel primo stadio, a partire dall'insieme reale di tutte le alternative, che si ricorda essere sconosciuto per dimensioni e caratteristiche sia al ricercatore che al viaggiatore, attraverso l'applicazione di una tecnica di generazione ed enumerazione dei percorsi si arriva alla costruzione dell'insieme principale. A questo livello, la combinazione delle caratteristiche individuali e dei tratti latenti che li contraddistinguono viene trattata attraverso un'euristica

di semplificazione (fase non-compensativa del processo), arrivando alla formazione dell'insieme di considerazione (Kaplan & Prato, 2012). Si entra quindi nel secondo stadio in cui, attraverso l'applicazione di una tecnica compensativa come la massimizzazione dell'utilità, nella quale vengono elaborati tutti gli attributi del percorso, si genera la scelta finale.

In definitiva, per la modellazione della scelta del percorso occorre tenere bene in considerazione i seguenti aspetti:

1. Per ogni coppia OD sono possibili diversi set di scelta, distinti per composizione e dimensione, e dipendenti dall'individuo;
2. In una rete di trasporto i percorsi non sono distinguibili in maniera univoca. Infatti, diversi itinerari possono condividere uno o più tratti, rendendoli simili e causando l'impossibilità di essere etichettati e distinti senza ambiguità.
3. Capita spesso che le reti utilizzate come base dei modelli di scelta del percorso siano delle rappresentazioni semplificate di quelle reali, non contenenti quindi tutti i possibili itinerari tra una determinata coppia OD e, quindi, è possibile che vi siano delle lacune riguardanti la possibilità di riprodurre il percorso effettivamente scelto dall'utente. Inoltre, non esiste un set universale di percorsi definito, in quanto è un fattore che richiede l'applicazione di particolari metodi di enumerazione degli stessi che saranno propedeutici per la costruzione dei diversi insiemi di scelta degli individui;
4. Le elevate dimensioni della rete e il numero di percorsi possibili tra le diverse coppie OD si traduce spesso in tempi computazionali non efficienti, ragione per cui nella scelta del percorso non sono direttamente utilizzabili le procedure già sviluppate per la gestione di altri ambienti di scelta, come il modo di trasporto, un capo di abbigliamento, una marca, etc., anche per il fatto che queste alternative sono mutualmente relazionate e numerose (Bovy, 2009).

Modellare separatamente la costruzione dell'insieme di scelta e la scelta stessa ha diversi lati positivi, perché sia nella stima dei modelli che nella previsione fornisce ampie opportunità di controllare il processo da tutti i punti di vista (Bovy, 2009).

### **2.2.5. La specificazione matematica**

Nei paragrafi precedenti si è illustrato il processo di scelta sia dal punto di vista del viaggiatore che da quello del ricercatore. Il momento successivo è quindi quello di tradurre in formulazioni matematiche quanto si evince dalla teoria. La regola base generale da cui poi è stata sviluppata una gran parte dei modelli di scelta del percorso presenti in letteratura

deriva dal paradigma di Manski (Manski, 1977), che ha tradotto in una formula matematica il processo semi-compensativo a due stadi, esprimendo che la probabilità, per un individuo  $i$ , di scegliere l'alternativa  $r$  dal proprio set di scelta  $CS_i$  (chiamato anche consideration set) è data dalla seguente espressione:

$$P_i(r|US_i) = \sum_{CS_i \in PS_i} p_i(r|CS_i) p(CS_i|US_i) \quad (1)$$

Dove

$P_i(r|US_i)$  è la probabilità che il soggetto  $i$  scelga l'alternativa  $r$  dall'universal set  $US_i$  di tutte le alternative disponibili per  $i$ ;

$p_i(r|CS_i)$  è la probabilità condizionata che l'utente  $i$  possa scegliere l'alternativa  $r$  dal consideration set  $CS_i$  sottoinsieme di  $US_i$ ;

$p(CS_i|US_i)$  probabilità che  $CS_i$  sia il consideration set dell'utente  $i$  dato il proprio universal set  $US_i$ .

La probabilità totale  $P_i(r)$  è quindi la somma effettuata su tutti i consideration set divenuti elementi del cosiddetto power set  $PS_i$ , che è il set di tutti i possibili choice set non vuoti.

Un elemento essenziale di questo paradigma è che la maggior parte delle volte il choice set di un individuo non è conosciuto con certezza e non può essere osservato, quindi è un concetto latente che potrebbe essere modellato come un insieme casuale. Tuttavia, questo non è conveniente dal punto di vista del calcolo della probabilità totale sulla somma di tutti i  $CS_i$ , perché il loro numero cresce esponenzialmente con il numero di alternative in  $US_i$ . Quindi, nonostante l'interpretazione sia realistica, nella pratica non viene applicata, utilizzando invece un approccio deterministico. Ne deriva quindi che la probabilità di scelta è data dalla seguente espressione:

$$P_i(r|US_i) = p_i(r|CS_i) \subseteq US_i \quad (2)$$

In molti casi ed in particolare nella scelta del percorso, l'insieme di scelta è sicuramente un sottoinsieme di un qualche set universale ritenuto disponibile. L'espressione di Manski mostra inoltre come i set di scelta giochino un ruolo sia nella stima dei modelli di scelta che nella previsione della domanda. Nel caso di stima, i parametri della funzione probabilistica di scelta  $p_i(r|CS_i)$ , implementati ad esempio in una struttura logit, sono determinati dato un choice set  $CS_i$  ritenuto rilevante e date osservazioni  $p_i$  uguali a 0 o 1 della scelta. Nel caso di previsione della domanda, la probabilità di usare l'alternativa  $r$  è prevista dato un predetto choice set  $CS_i$  e una funzione di scelta stimata  $p_i(r|CS_i)$ . In caso di stima, invece, il pedice  $i$  si riferisce al singolo individuo, nel caso di previsione della domanda, nella maggior parte dei casi, si riferisce ad un gruppo di individui simili che viaggiano tra la stessa origine e destinazione assumendo lo stesso set di scelta collettivo (Bovy, 2009).

### 2.3. Generazione dell'insieme di scelta

Nelle pagine precedenti è stata illustrata l'importanza che l'insieme di scelta assume nel processo di modellazione della scelta del percorso. In questo paragrafo si descrivono brevemente le tecniche che sono state sviluppate per la generazione dell'insieme di scelta, partendo dalla decisione se sia preferibile utilizzare un insieme di scelta implicito o esplicito. Dal punto di vista concettuale, razionale ed empirico, dovrebbe essere preferibile utilizzare una generazione dei percorsi esplicita a monte della stima dei modelli (Prato, 2009). Come indicato nei paragrafi precedenti, la formazione del set di scelta è una fase solitamente non compensativa. Infatti, è un processo legato ai vincoli delle alternative, perché percorsi con caratteristiche non desiderabili sono scartati dagli utenti; dipende inoltre dalle preferenze dell'utente, che influenzano il modo con cui i percorsi vengono considerati (Horowitz & Louviere, 1995); è anche un processo che segue la logica del trial-and-error, in quanto la percezione delle caratteristiche dei percorsi è condizionata dall'utilizzo o meno del percorso e dalle informazioni a disposizione su di esso (Richardson, 1982). Invece, la scelta tra le opzioni considerate è puramente compensativa, perché viene effettuato un "trade-off" tra gli attributi delle alternative per massimizzare l'utilità. Inoltre, la modellazione esplicita del set di scelta, nel caso della scelta del percorso, è necessaria anche per altre ragioni, tra cui le dimensioni della rete, i differenti insiemi di scelta possibili tra le osservazioni e i reali benefici dal punto di vista dell'implementazione di diverse politiche decisionali (Bovy, 2009). Alcuni casi di studio mostrano empiricamente i benefici della generazione esplicita del set di scelta. Cascetta, et al. (1997) hanno dimostrato che soluzioni path-based applicate al problema dell'equilibrio stocastico dell'utente (SUE), utilizzando una generazione esplicita dei percorsi, producano risultati migliori rispetto agli algoritmi deterministici e stocastici tradizionali con enumerazione implicita. Bekhor & Toledo (2005) argomentano che, per correttezza, gli algoritmi link-based necessitano di set di scelta impliciti, perché da un punto di vista comportamentale i percorsi generati potrebbero essere irrealistici. Ad esempio, percorsi considerati efficienti che conducono l'utente vicino alla destinazione e lontano dall'origine (Dial, 1971; Maher, 1998; Dial, 2006) possono essere irrealistici se vi è un continuo movimento di ingresso-uscita da autostrade, mentre percorsi molto lunghi derivanti dalla soluzione al problema dell'assegnazione del traffico (Bell, 1995; Akamatsu, 1996) difficilmente sono scelti dagli utenti. (Bliemer & Taale, 2006) illustrano che, in caso di reti estese, la generazione del set di scelta a priori dei calcoli per l'assegnazione offre diversi vantaggi teorici e computazionali per la parte relativa all'assegnazione del traffico nel caso di previsione della domanda di viaggio.

Il secondo aspetto da considerare è lo scopo della generazione dei percorsi. Possono essere individuate tre motivazioni principali: 1) analisi delle alternative di viaggio per

determinarne la disponibilità, numero, caratteristiche, varietà e composizione; 2) stima di modelli di domanda disaggregati per scoprire parametri comportamentali delle funzioni di utilità a livello individuale, attraverso l'utilizzo di osservazioni di scelte reali; 3) previsione delle probabilità di scelta per determinare i volumi di traffico sugli archi della rete, usando modelli di scelta del percorso con parametri stimati. Dati questi tre scopi principali, si andranno a considerare i requisiti principali che i set generati devono avere.

Dalla prospettiva dell'analisi delle alternative di viaggio, le statistiche sulle caratteristiche del percorso e le misure di distribuzione di similarità forniscono una panoramica delle alternative disponibili agli utenti. Sfortunatamente, gli insiemi di scelta realmente considerati sono raramente osservabili (Hoogendoorn-Lanser, 2005), anche se queste rare osservazioni forniscono alcuni spunti di riflessione. Gli utenti non considerano tutte le alternative possibili, ma piuttosto limitano la disponibilità dei percorsi sulla base dei loro vincoli, preferenze ed esperienze. Ad esempio, alcuni percorsi possono non essere considerati a causa della presenza di circuiti, mentre altri possono non essere percepiti come alternative differenti a causa dell'elevata sovrapposizione. Ancora, alcuni percorsi potrebbero non essere soddisfacenti a causa di una combinazione di vincoli, mentre altri possono non essere possibili a causa di irrealistici passaggi ripetuti tra strade della stessa categoria funzionale (Prato, 2009). Dalla prospettiva della stima del modello di domanda, Bovy (2009) argomenta che i requisiti della qualità degli insiemi di scelta non sono restrittivi, così come non tutte le alternative rilevanti devono essere incluse, perché è possibile ottenere risultati soddisfacenti in termini di stima dei parametri di utilità anche quando sono utilizzati insiemi di scelta di ridotte dimensioni. Tuttavia, Bekhor & Prato (2006), Prato & Bekhor (2007) e Bliemer & Bovy (2008) illustrano che la dimensione e composizione dell'insieme di scelta influiscono sulle stime e sulle probabilità di scelta per alcune specificazioni modellistiche. Apparentemente, l'argomento proposto da Bovy (2009) è corretto solo quando è stimato il multinomial logit (MNL) o le sue modificazioni, non quando l'eterogeneità dei gusti, schemi di sostituzione o eteroschedasticità sono introdotti in modelli GEV (Generalized Extreme Value) o Mixed (Prato, 2009).

Dal punto di vista della previsione dei percorsi e dei volumi di traffico sugli archi, Bovy (2009) dice che i requisiti di qualità dell'insieme di scelta sono molto restrittivi, così come è necessario che tutte le alternative rilevanti vengano incluse al suo interno. Aggiunge inoltre che l'inclusione nell'insieme di scelta di percorsi non attrattivi non dovrebbe distorcere le previsioni di domanda e nemmeno influenzare l'efficienza computazionale. Tuttavia, una definizione obiettiva dei percorsi rilevanti nelle reti reali è ancora mancante, quindi la valutazione sulla correttezza dell'insieme di scelta usato in caso di previsione appare legata alla sensibilità ed esperienza dell'analista, piuttosto che da obiettive misure di qualità dello stesso.

A partire da queste premesse, nel seguito saranno brevemente descritte le tecniche più comunemente usate per la generazione esplicita dell'insieme di scelta. È importante sottolineare che non solo l'insieme universale sia estremamente vasto e complesso, ma anche i set di scelta generati possono essere facilmente di grandi dimensioni, eterogenei in termini di composizione dei percorsi, complessi riguardo la sovrapposizione fisica tra gli stessi e dipendenti dagli schemi di caricamento che influenzano i valori degli attributi di percorso. La classificazione dei metodi di generazione dell'insieme di scelta si divide in quattro categorie: metodi di ricerca del percorso minimo deterministici o stocastici, algoritmi a enumerazione vincolata e approcci probabilistici.

### **2.3.1. Metodi deterministici**

Il gruppo più ampio di metodologie per la generazione dei percorsi è basato sul calcolo del percorso più breve, che avviene attraverso la modifica di una o più variabili di input della rete, come impedenze sugli archi, vincoli di percorso, criteri di ricerca. Una caratteristica comune di queste tecniche è che i risultati sono deterministici, le coppie origine-destinazione sono processate sequenzialmente e la maggior parte delle procedure sono euristiche (fatta eccezione per gli algoritmi di ricerca dei K percorsi più brevi).

#### **Algoritmi di calcolo del percorso più breve**

L'approccio più semplice alla soluzione del problema della generazione dell'insieme di scelta consiste nel calcolo dei migliori K percorsi in rispetto di funzioni di costo generalizzato additive sugli archi. L'assunzione comportamentale alla base della ricerca dei percorsi è che i viaggiatori limitano le loro scelte attorno ad un certo numero di percorsi di minimo costo ed evitano la scelta di alternative estremamente costose. Il numero di percorsi generati definisce una finestra di costi ammissibili tra il percorso più breve e il k-esimo percorso più breve, così come un numero elevato porta ad una maggiore disponibilità di alternative e includerà anche alternative molto più lunghe e costose rispetto al reale percorso più breve. Alcuni algoritmi esistenti estendono l'approccio adottato per il calcolo del percorso più breve (Bellman, 1958; Dijkstra, 1959). In letteratura sono presenti anche metodi che offrono una soluzione esatta al problema del k-esimo percorso più breve. Alcuni algoritmi permettono cicli all'interno del percorso (Hoffman & Pavley, 1959; Bellman & Kalaba, 1960; Dreyfus, 1969; Shier, 1976; Shier, 1979; Eppstein, 1998), mentre altri algoritmi considerano solamente percorsi aciclici (Yen, 1971; Lawler, 1976; Katoh, et al., 1982; Hadjiconstantinou & Christofides, 1999).

Sia per finalità di stima che di previsione, il limite della tecnica di calcolo dei K percorsi più brevi è quella di generare percorsi eccessivamente tortuosi e estremamente simili che non risultano attrattivi. Una possibile soluzione consiste nell'implementazione di uno degli algoritmi di ricerca del k-esimo percorso minimo che consideri itinerari aciclici con un elevato numero di percorsi generati. Bekhor, et al. (2006) usa l'algoritmo per il calcolo dei K percorsi più brevi con K pari a 15 e 40, confrontando i percorsi generati con quelli osservati (registrati tra il personale del M.I.T). Sorprendentemente, l'aumento percentuale dei percorsi riprodotti è pari al 1% quando sono aggiunti 25 percorsi. Inoltre, il 30% dei percorsi osservati non sono replicati dall'algoritmo, mostrando quindi come quest'ultimo produca diverse alternative irrilevanti. L'aumento del numero dei K percorsi sembra quindi non risolvere il problema dell'elevata similarità, altrimenti la percentuale di riproduzione dovrebbe crescere maggiormente all'aumentare di K.

In letteratura sono state proposte delle soluzioni alternative per tenere conto delle preferenze degli individui, attraverso l'incremento dell'eterogeneità dei percorsi. Lombard & Church (1993) introducono il concetto di "gateway shortest path", vincolando i percorsi attraverso nodi (gateways) sparsi nella rete per permettere una generazione dei percorsi maggiormente distribuita nello spazio. Kuby, et al. (1997) propongono un metodo iterativo che seleziona un sottoinsieme di K percorsi più brevi minimizzando una misura di similarità basata sugli archi condivisi tra i percorsi. Hunt & Kornhauser (1997) presentano il concetto di "essentially least cost path" lavorando su una soglia di costo per costruire percorsi che contengano accettabili deviazioni locali dal quello di minimo costo. Akgün, et al. (2000) trattano l'enumerazione dei percorsi come un problema di "dispersione p" per generare percorsi dissimili tra loro, applicando un'euristica che trova una soluzione iniziale dei k percorsi più brevi e quindi opera una ricerca locale basata sulla minimizzazione della similarità, in modo tale da migliorare questa soluzione iniziale. Van der Zij & Fiorenzo-Catalano (2005) propongono un metodo alternativo per trovare i percorsi possibili in accordo con una vasta classe di vincoli comportamentali, mantenendo allo stesso tempo l'efficienza computazionale degli altri algoritmi.

Queste tecniche generano complessivamente percorsi aciclici ed eterogenei. Tuttavia, l'implementazione suggerisce due limitazioni in caso rispettivamente di stima e previsione. Dal punto di vista modellistico, tutti i viaggiatori che si muovono tra la stessa origine e destinazione condividono lo stesso insieme di scelta generato, ma questa proposizione appare irrealistica perché ci dovrebbe essere un'assunzione comportamentale volta ad assicurare che ogni utente abbia personali vincoli, preferenze ed esperienze che lo portino ad avere un proprio insieme di scelta. Dal punto di vista della previsione, l'assenza di misure di attrattività oggettive rimanda all'esperienza e sensibilità del ricercatore il controllo sulla qualità attrattiva dei percorsi generati (Prato, 2009).



### **Labeling approach**

L'assunzione comportamentale alla base del labeling approach (letteralmente: etichettatura) è quella che i viaggiatori hanno diversi obiettivi. Alcuni preferiscono minimizzare il tempo di viaggio, altri magari non si sentono a loro agio in strade e autostrade pesantemente congestionate. Qualcuno preferisce viaggiare attraverso punti di riferimento familiari, mentre altri potrebbero optare per percorsi più panoramici in assenza di vincoli temporali. Ogni criterio corrisponde ad un differente percorso preferito, e ognuno di questi dovrebbe essere etichettato in funzione del differente obiettivo per il quale è stato scelto come ottimale.

Ben-Akiva, et al. (1984) propongono un approccio di questo tipo nel quale i percorsi fisici compongono il "nest" superiore di una struttura modellistica nested logit (NL). Siccome i viaggiatori scelgono tra i percorsi fisici e non tra le etichette, il settaggio per la stima del NL è modificato rispetto a quello tradizionale dove invece è conosciuta la scelta al livello (nest) più basso. La stima di un modello di questo tipo consiste nel calcolo analitico dei valori inclusivi e nella stima della massima verosimiglianza del modello NL non lineare nei parametri. Alcune specificazioni considerano inoltre attributi di percorso e combinazione di etichette come "dummies" nel nido (nest) dei percorsi fisici, con solo costanti specifiche dell'etichetta al livello del percorso etichettato.

Ramming (2002) presenta una versione semplificata dell'approccio e cerca il percorso più breve sulla base di 16 etichette, partendo dalla regola più convenzionale della ricerca del minimo tempo di viaggio a quella meno convenzionale di massimizzazione del tempo di viaggio trascorso in zone sicure. Prato & Bekhor, 2006) (2006) considerano quattro attributi e confrontano i percorsi creati rispetto a quelli rilevati attraverso un'indagine che ha coinvolto la facoltà e il personale del Politecnico di Torino. Entrambi i casi di studio mostrano che il labeling approach riproduce solo parzialmente le scelte reali degli individui, suggerendo inoltre che gli indici di sovrapposizione non soddisfacenti dipendano dalla definizione delle etichette. La discrezione dell'analisi guida questa metodologia, e la corretta interpretazione della funzione obiettivo per ciascun individuo è un problema complesso che, per un'implementazione efficiente, richiede una conoscenza a priori delle preferenze individuali.

### **Link elimination**

Questo approccio è basato sulla ricerca ripetitiva dei percorsi di minimo costo dopo la rimozione di parte o dell'intero set di archi che sono già stati scelti nelle iterazioni precedenti. Dal punto di vista comportamentale questo approccio garantisce non similarità

tra le diverse alternative che, ovviamente, non condividono archi nella misura definita dalla regola di eliminazione. Queste regole sono lasciate alla libera interpretazione del ricercatore, che può per esempio eliminare tutti gli archi del percorso realmente più breve, un arco che porta l'utente lontano dalla destinazione e vicino all'origine, oppure un arco che appartiene a strade di categoria bassa.

Bellman & Kalaba (1960) illustrano un'euristica simile all'approccio ad eliminazione. La procedura di "ramificazione" (branching) calcola il percorso più breve, identifica gli archi che hanno un nodo che appartiene al percorso e l'altro no, e selezionano un arco che soddisfa una predefinita regola euristica (ad esempio il primo arco che comincia dall'origine, o un arco che ha la massima capacità). Un nuovo percorso è composto calcolando il percorso più breve dalla testa o dalla coda dell'arco selezionato fino alla destinazione. Utilizzando altri archi possono essere selezionati percorsi aggiuntivi. Azevedo, et al. (1993) presentano l'approccio ad eliminazione originale nel quale la regola di rimozione elimina tutti gli archi del percorso più breve prima di calcolarne un altro. Il difetto di questa metodologia è la disconnessione della rete, perché la rimozione di centroidi connettori e dei principali attraversamenti non garantisce l'esistenza di percorsi addizionali tra l'origine e destinazione dello spostamento (Prato, 2009).

Bekhor, et al. (2006), Prato & Bekhor (2006) e Frejinger & Bierlaire (2007) applicano una variazione all'approccio originale, calcolando il percorso di tempo minore rimuovendo un arco dal percorso più breve ad ogni 50, 10 e 50 iterazioni, rispettivamente. Solo Prato e Bekhor (2006) spiegano nel dettaglio la procedura adottata, che rimuove un arco quando il nodo finale è più lontano dalla destinazione e più vicino all'origine rispetto a quello iniziale. Quando sono confrontati con le scelte reali degli utenti, questi tre studi riportano una percentuale di riproduzione che varia tra il 60% e il 80%.

L'implementazione della link elimination appare problematica da un punto di vista comportamentale. Quando è utilizzata la formulazione originale, i problemi di disconnessione della rete non consentono la generazione di tutti i percorsi attrattivi usando accessi alternativi a quelli principali, anche se gli utenti spesso li utilizzano. Quando è applicata la singola eliminazione, il metodo produce percorsi estremamente simili, perché per ogni rimozione di un arco esiste una piccola deviazione che utilizza archi vicini a quello eliminato. L'implementazione della variante dell'approccio originale sembra problematica anche dal punto di vista computazionale, perché percorsi realistici possono includere un elevato numero di archi e il numero di combinazioni possibili di archi eliminabili possono scoraggiare la rimozione di un arco singolo (Prato, 2009).

## **Link Penalty**

Anche questo approccio, come i precedenti, è basato sulla ricerca ripetitiva dei percorsi di minimo costo, ma è imposta una penalità nell'impedenza di tutti gli archi del percorso risultante più breve, invece che la rimozione dell'arco. Dal punto di vista comportamentale, anche questo metodo scoraggia la selezione dello stesso insieme di archi e prova a generare percorsi che sono più diversi all'aumentare della penalizzazione.

De la barra et al. (1993) introduce l'euristica di penalizzazione dell'arco, che costruisce l'insieme di scelta immagazzinando iterativamente i percorsi. Il percorso salvato per primo è quello realmente più breve, mentre quelli successivi sono quelli calcolati dopo l'applicazione di un aumento percentuale fissato delle impedenze sugli archi appartenenti al percorso precedente. La procedura termina quando due percorsi salvati coincidono. Ruphail e al. (1995) definiscono il metodo di penalizzazione iterativo (IPM - Iterative Penalty Method), dove la penalizzazione può essere applicata non solo al percorso correntemente più breve, ma anche a tutti gli altri che sono stati calcolati, aggiungendo un'impedenza fissa oppure moltiplicando quelle fisse per un termine prefissato. Park and Rilett (1997) propongono una variante all'approccio originale per evitare la generazione di deviazioni minori all'inizio e alla fine del percorso. Nello specifico, le impedenze degli archi non sono incrementate all'interno di una certa distanza dall'origine o destinazione dello spostamento. Di conseguenza, i percorsi generati sono meno simili e aumenta la rilevanza delle alternative. Scott et al. (1997) arricchiscono l'approccio con un programma di ottimizzazione per determinare il fattore di penalità dei percorsi più brevi, in modo da generare il successivo con una sovrapposizione rispetto all'ultimo non superiore a un determinato numero di archi.

Bekhor et al. (2006) definiscono i fattori di penalità proporzionalmente alla distanza tra l'origine e la destinazione dei percorsi osservati, per cui vi è un incremento maggiore per distanze più lunghe. Prato e Bekhor (2006) identificano un unico fattore di penalità e compiono 15 iterazioni. Entrambi gli studi mostrano che la penalità sull'arco non fornisce un miglioramento in termini di generazione di percorsi rilevanti, poiché il metodo riproduce percentuali leggermente inferiori di percorsi osservati rispetto all'approccio con eliminazione degli archi.

I vantaggi della link penalty sono dati dal fatto che gli archi fondamentali rimangono nella rete e dallo "scoraggiamento" degli archi già individuati ad appartenere ai futuri percorsi calcolati. I limiti di questo approccio consistono nella generazione di percorsi con forti impedenze che indeboliscono la rilevanza dei percorsi e nella dipendenza della definizione del fattore di penalità. Infatti, utilizzando valori bassi l'algoritmo non è efficiente dal punto di vista computazionale, perché lo stesso percorso è identificato ripetutamente, mentre valori

alti portano alla generazione di percorsi con alte impedenze e poco attrattivi prima di quelli più brevi e maggiormente allettanti (Prato, 2009).

### **2.3.2. Metodi stocastici**

Un piccolo gruppo di metodi di generazione dei percorsi è basato sulla ricerca ripetuta del percorso di minimo costo, dove il calcolo di quello ottimale è effettuato attraverso un'estrazione casuale, da una specifica distribuzione di probabilità, delle impedenze sugli archi e delle preferenze individuali. Caratteristiche comuni di questi metodi sono che i risultati sono stocastici, le coppie origine-destinazione sono processate simultaneamente e tutte le procedure sono euristiche. Oltretutto, la generazione stocastica è un caso di "importance sampling", perché la probabilità di selezione di un percorso dipende dalle proprietà dello stesso, come la lunghezza o il tempo di viaggio. Di conseguenza, i modelli di scelta del percorso basati su questi metodi necessitano di un termine di correzione per il campionamento che tenga conto delle diverse probabilità di selezione tra i percorsi alternativi, in modo da non produrre delle stime dei parametri errate (Frejinger, 2007; Bovy, et al., 2009; Frejinger, et al., 2009).

### **Approcci simulativi**

L'assunzione comportamentale alla base dell'approccio simulativo è che i viaggiatori percepiscono i costi di percorso con un errore, che è rappresentato dall'estrazione di funzioni di costo generalizzato da distribuzioni probabilistiche.

L'idea dell'approccio simulativo ha origine dall'applicazione del modello Multinomial Probit (MNP) all'assegnazione del traffico. Sheffi e Powell (1982) sviluppano una procedura che, per ogni iterazione, implementa una tecnica Monte Carlo per estrarre i tempi di viaggio da una distribuzione Probit nell'intorno della funzione di costo complessivo congestionato. Di seguito, è compiuta una assegnazione "Tutto o Niente" e vengono calcolati i flussi finali sugli archi come i valori medi di quelli delle iterazioni precedenti. Cascetta (2001) argomenta che la simulazione della distribuzione dovrebbe essere effettuata solo nell'intorno della parte non congestionata della funzione di costo. Fiorenzo-Catalano e Van der Zijpp (2001) implementano una versione della tecnica Monte Carlo aumentando gradualmente la varianza dei componenti random del modello, in modo tale da mantenere ad un tasso costante la frequenza con cui sono trovati nuovi percorsi. Ramming (2002) propone la replica della stessa procedura per la generazione dei percorsi, ma ad ogni iterazione, invece del flusso sugli archi, viene salvato il percorso più breve.

Nielsen (2000) effettua delle estrazioni random da una distribuzione normale troncata in uno studio riguardante il progetto della ferrovia Copenhagen-Ringsted. Ramming (2002) effettua 48 estrazioni da una distribuzione normale, con media e varianza uguali al tempo di viaggio. Bierlaire e Frejinger (2005) ne effettuano 20 da una normale troncata, con media e varianza uguali al tempo di viaggio registrato da dispositivi GPS durante uno studio di sicurezza del traffico nella città di Borlange, Svezia. Prato e Bekhor (2006) effettuano 35 estrazioni da una normale troncata con media e varianza uguali al tempo di viaggio, con troncatura sinistra limitata al tempo di viaggio a flusso libero e troncatura destra uguale a quello calcolato assumendo una velocità di percorrenza di 10 km/h.

I vantaggi di questo metodo sono legati alla generazione di un largo numero di percorsi attrattivi, dipendenti dalla selezione di un'adeguata distribuzione probabilistica e numero di estrazioni. Se garantita un'adeguata correzione per le diverse probabilità di selezione (Frejinger, 2007; Bovy, et al., 2009; Frejinger, et al., 2009), gli insiemi di scelta generati stocasticamente sono utilizzabili sia per le stime modellistiche che per prevedere i flussi, condizionalmente alla definizione di un elevato numero di percorsi che dovrebbe contenere tutte le alternative rilevanti.

Le limitazioni del metodo sono legate alla discrezione nella scelta della distribuzione probabilistica e nel numero di estrazioni. La normale potrebbe rappresentare la distribuzione dei costi percepiti, ma i valori negativi dovrebbero essere troncati, perché le impedenze sugli archi non possono essere negative. Diversamente dalla normale, la normale troncata non è additiva nella media e nella varianza. Nell'applicazione della simulazione nel problema SUE, questa proprietà non garantisce l'indipendenza dalla segmentazione degli archi (Nielsen, 2000), e le distribuzioni come la log-normale e la gamma sono preferite perché forniscono solo valori positivi (Nielsen & Frederiksen, 2006). Nell'applicazione della simulazione in caso di singola estrazione casuale, la distribuzione non deve necessariamente essere riprodotiva e le normali troncate sono ammesse, anche se Nielsen (2000) dice che la troncatura porta a errori verso certi percorsi e quindi è preferita la distribuzione gamma.

Ramming (2002) effettua un numero di estrazioni che contiene il peso computazionale all'interno dei confini di quelli della link elimination e link penalty, perché i tempi di calcolo della simulazione sono proporzionali al numero di estrazioni. Si noti che l'incremento del numero di estrazioni aumenta sensibilmente i tempi computazionali ma non garantisce la generazione di ulteriori percorsi unici. Prato e Bekhor (2006) effettuano un numero di estrazioni che soddisfano l'abilità del metodo nel generare percorsi unici, perché la deviazione standard limita la possibilità di generare percorsi molto diversi. L'algoritmo non è invece efficiente con bassi valori della varianza, perché riproduce solo alcuni percorsi unici,

mentre valori alti non sono efficienti perché portano alla generazione di tanti percorsi irrealistici (Prato, 2009).

### **Approcci doppiamente stocastici**

L'assunzione comportamentale alla base di questi approcci è che non solo i viaggiatori percepiscano con errore i costi di percorso, ma differenti utenti abbiano diverse percezioni. Conseguentemente, la funzione di generazione ha un termine casuale legato al costo generalizzato e un altro per l'eterogeneità delle preferenze.

Nielsen (2000) introduce l'idea della funzione doppiamente stocastica sviluppando una modifica euristica al problema SUE con un primo componente che considera le percezioni dell'utente a livello di singolo arco, e una seconda che tiene conto delle differenze al livello delle funzioni di utilità dei singoli viaggiatori. Bovy e Fiorenzo-Catalano (2007) propongono di utilizzare una funzione di utilità dello spostamento come base per una funzione di generazione doppiamente stocastica. I percorsi rilevanti sono creati attraverso la ricerca di quello ottimale, variando stocasticamente gli attributi della rete e di preferenza. La variazione delle impedenze sugli archi riflette le differenze tra le preferenze dei viaggiatori riguardo questi attributi.

Bliemer e Taale (2006) applicano la funzione di generazione alla rete principale del modello nazionale tedesco, a priori del calcolo delle previsioni dei volumi di traffico con un modello di assegnazione dinamica. Bovy e Fiorenzo-Catalano (2007) calibrano la varianza e applicano l'approccio al corridoio Rotterdam-Dordrecht, in Olanda. La calibrazione dei parametri della funzione coinvolge un insieme di valori di varianza che devono essere determinati massimizzando alcune misure di somiglianza con le osservazioni degli insiemi di considerazione individuali o sui percorsi scelti.

I vantaggi del metodo sono l'eterogeneità delle alternative generate, la rilevanza dei percorsi data l'elevata conformità con quelli osservati, e l'efficienza computazionali in reti estese. La debolezza della tecnica consiste nella calibrazione dei coefficienti della funzione di probabilità, perché gli insiemi di considerazione sono difficili da osservare e l'uso di valori scorretti potrebbe portare alla generazione di percorsi irrealistici e irrilevanti. Inoltre, questa tecnica affronta gli stessi problemi dell'approccio simulativo in relazione alla selezione delle distribuzioni di probabilità, con il problema aggiuntivo della distribuzione del valore del tempo. Si noti che quando si considerano contemporaneamente costi e tempi, una distribuzione normale dei coefficienti di costo implica un'inaccettabile distribuzione del valore del tempo con media e varianza indefinite (Prato, 2009).

### 2.3.3. Metodi ad enumerazione vincolata

Questi metodi dipendono dall'assunzione che gli individui scelgano i percorsi in accordo con regole comportamentali oltre che al percorso di minimo costo.

Prato e Bekhor (2006) propongono un algoritmo di "branch and bound" dove la regola di ramificazione riflette le assunzioni comportamentali attraverso la definizione di soglie. Una soglia direzionale esclude la considerazione di archi che portano gli utenti significativamente lontano dalla destinazione e più vicino all'origine dello spostamento. Una soglia temporale consente il rifiuto di percorsi che gli utenti potrebbero considerare irrealistici a causa dell'elevato tempo di viaggio. Una soglia di circuitazione scarta gli archi che i viaggiatori non dovrebbero considerare perché includono lunghe deviazioni. Una soglia di similarità rimuove i percorsi con grandi tratti in sovrapposizione che gli utenti non considererebbero come alternative diverse. Una soglia di vincolo elimina invece i percorsi irrealistici che includono manovre che causano elevati ritardi in termini di tempo e apprensione negli autisti in approccio all'intersezione (Prato & Bekhor, 2006). L'algoritmo costruisce l'albero di connessione tra l'origine e destinazione dello spostamento processando sequenze di archi in accordo con la regola di ramificazione che tiene conto dei vincoli logici per aumentare l'eterogeneità dei percorsi generati. Ogni sequenza di archi che connette le origini e destinazioni e soddisfa tutti i vincoli viene inserita nell'insieme di scelta come soluzione possibile al problema della generazione dei percorsi (Prato & Bekhor, 2006). L'algoritmo elabora precedenti applicazioni di tecniche di enumerazione legate ai trasporti. Friedrich et al. (2001) applicano una procedura di assegnazione branch and bound su reti di trasporto pubblico usando un algoritmo di ricerca basato sugli orari. Hoogendoorn-Lanser (2005) adattano la stessa procedura per la generazione dell'insieme di scelta in un caso di rete di trasporto multimodale. In queste applicazioni la tecnica sfrutta sezioni di percorso predefinite, un'assunzione molto restrittiva per la generazione dei percorsi.

Prato e Bekhor (2006) mostrano che l'algoritmo produce percorsi eterogenei e rilevanti, perché sono riprodotte quasi tutte le scelte osservate. Bekhor e Prato (2009) applicano l'algoritmo al caso di studio di Boston e confermano che le assunzioni comportamentali, diversamente da quella della selezione del percorso di costo minimo, sono più efficienti e producono percorsi realistici. Da un punto di vista computazionale, la velocità dell'algoritmo dipende esponenzialmente dal dettaglio dell'albero e conseguentemente dal numero di archi del percorso. L'implementazione dell'algoritmo, di conseguenza, è consigliata su reti di modesta dimensione (vedere ad esempio Frejinger, 2007), anche se i risultati empirici provano l'opposto quando sono utilizzate diverse strutture dei parametri nella rete estesa di Boston rispetto a quella piccola di Torino (Bekhor & Prato, 2009).

Dal punto di vista della stima, i metodi ad enumerazione vincolata producono un insieme di scelta esaustivo, estremamente utile per la stima dei parametri di utilità. Dalla prospettiva della previsione, il problema principale rimane quello della difficoltà di determinare se sono stati generati tutti i percorsi rilevanti. Tuttavia, la natura esaustiva ed esatta della soluzione suggerisce che saranno generati tutti i percorsi attrattivi insieme ad alcuni che non lo sono, senza problemi significativi per la previsione dei flussi (vedere Bovy, 2009). La debolezza del metodo è legata alla definizione delle soglie dei vincoli comportamentali. L'assenza di un'analisi di sensitività riflette la mancanza di conoscenza sugli effetti della variazione delle soglie comportamentali (Prato, 2009).

#### **2.3.4. Metodi probabilistici**

Questi metodi associano una probabilità di generazione ad ogni percorso. L'adozione dell'approccio interamente probabilistico di Manski (1977) è impossibile nella scelta del percorso. Infatti, anche per un piccolo insieme di scelta di dieci percorsi (piccolo in caso di reti urbane dense), sembra impraticabile e poco conveniente calcolare la probabilità di ciascuno dei potenziali 1023 insiemi di scelta.

Cascetta e Papola (2001) propongono il modello a percezione/disponibilità implicita (Implicit Availability/Perception - IAP) dove la probabilità di un'alternativa di appartenere all'insieme di scelta si inserisce nella funzione di utilità del modello di scelta. Intuitivamente, una bassa probabilità di appartenenza porta ad una bassa probabilità di scelta dell'alternativa. Anche se la probabilità di appartenenza è supposta dipendere dalle attitudini e dalla percezione delle variabili da parte degli individui, l'unica stima di questo modello andata a buon fine mostra che questa probabilità dipenda da variabili socio-economiche e dagli attributi di utilità. Ramming (2002) descrive nel dettaglio come qualsiasi tentativo di stimare lo IAP Logit con variabili legate alla conoscenza della rete non produce risultati soddisfacenti.

Frejinger (2007) e Frejinger et al. (2009) calcolano la probabilità per ogni arco della rete. Data una coppia origine-destinazione, la probabilità dell'arco dipende dalla sua distanza dal percorso di costo minimo in accordo ad una funzione di costo generalizzato. Di conseguenza, tutti gli archi nel reale percorso più breve avranno probabilità uguale a uno, mentre tutti gli altri tra zero e uno. Partendo dall'origine, una procedura ripetuta di random walk (cammino casuale) aggiunge archi successivamente da nodo a nodo, con il processo di selezione dell'arco ad ogni nodo che è governato dalla probabilità associata all'arco successivo. Alla destinazione, la probabilità del percorso corrisponde al prodotto di quelle associate agli archi che lo compongono, ed è utilizzata per correggere dall'errore dovuto al



campionamento quando l'insieme di scelta risultante è usato per le stime dei modelli. L'applicazione della metodologia avviene su una rete sintetica per illustrare i vantaggi del protocollo di generazione. I risultati mostrano che i modelli che includono la correzione per il campionamento sono molto migliori rispetto a quelli che ne sono sprovvisti. Tuttavia, i risultati positivi di questa applicazione sembrano legati alla inusuale rete sintetica, disegnata in maniera tale da evitare percorsi lunghi e ciclici, piuttosto che alla bontà del metodo. Infatti, i percorsi generati con il random walk potrebbero essere particolarmente circuitosi, contenere cicli ed essere estremamente lunghi fino a non raggiungere la destinazione in numero di iterazioni ragionevole, diventando quindi poco allettante dal punto di vista della stima e previsione (Prato, 2009).

#### **2.4. Modelli di scelta del percorso**

I sottoinsiemi di alternative costruiti con le tecniche di generazione descritte in precedenza sono generalmente abbastanza vasti sia per la stima dei modelli che per la previsione dei flussi, perché sono presenti, molto probabilmente, tutti i percorsi rilevanti e anche alcuni altri che non lo sono. Casi di studio su reti urbane americane ed europee hanno mostrato che per la stima dei modelli sono maggiormente utilizzate fino a 70 alternative, mentre 100 costituiscono la pratica comune adottata in caso di assegnazione del traffico. Intuitivamente, nel contesto della scelta del percorso, il numero delle alternative nell'insieme di scelta gioca un ruolo fondamentale nella stima di modelli di scelta discreta.

Recenti studi forniscono alcune idee sugli effetti dell'insieme di scelta nelle stime modellistiche e sulle prestazioni. Bekhor et al. (2006) esaminano gli effetti della dimensione del campione in assegnazioni path-based (basate sui percorsi) attraverso l'analisi della rete di Winnipeg, mostrando che insiemi di scelta più grandi migliorino la convergenza e i valori delle funzioni obiettivo. Prato e Bekhor (2007) illustrano una significativa influenza della dimensione e composizione degli insiemi dei percorsi generati sulle stime dei parametri e sulle performance del modello, anche se i risultati mostrati sono riportati in forma aggregata e non si riferiscono al singolo insieme di scelta. Frejinger (2007) mostra che l'insieme di scelta completo è necessario per calcolare le misure di correlazione e ottenere risultati di stima ottimali. Bliemer e Bovy (2008) affrontano il problema della previsione della scelta del percorso e mostrano come l'aggiunta di alternative irrilevanti causa errori nelle probabilità di scelta e fa sì che percorsi attrattivi diventino meno appetibili, in contrasto con la robustezza del modello rispetto all'aggiunta di percorsi irrilevanti (Bovy, 2009).

In accordo con quanto detto finora, i modelli di scelta del percorso dovrebbero mostrare robustezza nelle stime dei parametri di utilità rispetto alla dimensione dell'insieme di scelta.

In caso di stima, questi requisiti dovrebbero consentire la definizione di insiemi di scelta con un ragionevole numero di percorsi attrattivi, in modo da ottenere stime dei parametri affidabili. In caso di previsione, questi requisiti modellistici potrebbero non risolvere il problema della necessità di generare tutti i possibili percorsi rilevanti (Prato, 2009).

Reti urbane dense con 70 o 100 alternative mostrano un alto grado di similarità tra i diversi percorsi. Per questa ragione, la maggior parte della letteratura si concentra sulla correlazione tra le alternative, che altera le probabilità di scelta di percorsi che hanno sovrapposizione. Prashker e Bekhor (2004) illustrano il semplice problema di sovrapposizione, che è la scelta tra il classico autobus rosso e quello blu, ed un caso di cambio percorso per osservare che alcuni modelli di scelta del percorso non catturino correttamente la similarità tra le alternative in ciascuna configurazione della rete. Bliemer e Bovy (2008) utilizzano una simulazione Monte Carlo di scelte del percorso da una rete a griglia per mostrare che percorsi con la stessa lunghezza non sono preferiti allo stesso modo, perché al crescere della sovrapposizione diminuisce la propria probabilità in favore di alternative più indipendenti. Chiaramente, i modelli di scelta del percorso dovrebbero essere capaci di rappresentare in maniera adeguata la struttura di correlazione tra le alternative. Non è certo che la similarità sia percepita a livello di arco o a quello di gerarchizzazione stradale, e i modelli con la capacità di integrare entrambi questi aspetti sarebbero da preferire. Si noti che i modelli di scelta discreta più comuni nella pratica della modellazione del comportamento di viaggio, MNL e NL, non sono modelli di scelta del percorso adatti. Il MNL non consente di tenere conto della similarità delle alternative, mentre il NL assume che ogni alternativa appartenga esclusivamente ad un nido (nest), mentre nelle reti reali centinaia di percorsi condividono gli stessi archi. L'ultimo aspetto da trattare riguarda la previsione dei flussi di arco e percorso. Le formulazioni SUE sono disponibili per modelli Logit e GEV (Generalized Extreme Value), non per modelli più potenti che tengono conto di strutture di similarità complesse e potenzialmente per eterogeneità delle preferenze e eteroschedasticità. I modelli di scelta del percorso dovrebbero poter essere tradotti in equivalenti formulazioni matematiche di un problema di assegnazione del traffico per scopi di previsione. La classificazione di differenti modelli di scelta del percorso dipende quindi dalle differenti strutture modellistiche. La loro valutazione dipende dalla soddisfazione dei sopracitati requisiti e dai risultati di casi di studio su reti reali presenti in letteratura (Prato, 2009).

### 2.4.1. Modelli Logit

I modelli MNL modificati mantengono la stessa struttura logit, introducendo un fattore di correzione nella componente deterministica della funzione di utilità per approssimare la correlazione tra i diversi percorsi.

#### C-Logit

Cascetta et al. (1996) propongono una modifica del modello MNL, dove un fattore di comunanza misura il grado di similarità di ogni percorso con gli altri appartenenti all'insieme di scelta  $C$ . L'espressione della probabilità  $P_k$  di scegliere il percorso  $k$  dall'insieme di scelta  $C$  riflette la semplice struttura logit del modello:

$$P_k = \frac{\exp(V_k + \beta_{CF} \cdot CF_k)}{\sum_{l \in C} \exp(V_l + \beta_{CF} \cdot CF_l)} \quad (3)$$

dove  $V_k$  e  $V_l$  sono rispettivamente le funzioni di utilità del percorso  $k$  e  $l$ ,  $CF_k$  e  $CF_l$  sono i fattori di comunanza e  $\beta_{CF}$  è un parametro da stimare. In letteratura sono presenti diverse formulazioni del fattore di comunanza (Cascetta, et al., 1996; Cascetta & Papola, 2001):

$$CF_k = \ln \sum_{l \in C} \left( \frac{L_{kl}}{\sqrt{L_k L_l}} \right)^{\gamma_{CF}} \quad (4)$$

$$CF_k = \ln \sum_{a \in \Gamma_k} \left( \frac{L_a}{L_k} \sum_{l \in C} \delta_{al} \right)^{\gamma_{CF}} \quad (5)$$

$$CF_k = \ln \sum_{a \in \Gamma_k} \left( \frac{L_a}{L_k} \ln \sum_{l \in C} \delta_{al} \right)^{\gamma_{CF}} \quad (6)$$

$$CF_k = \ln \left[ 1 + \sum_{\substack{l \in C \\ k \neq l}} \left( \frac{L_{kl}}{\sqrt{L_k L_l}} \right) \left( \frac{L_k - L_{kl}}{L_l - L_{kl}} \right) \right] \quad (7)$$

dove  $L_k$  e  $L_l$  sono rispettivamente le lunghezze dei percorsi  $k$  e  $l$ ,  $L_a$  è la lunghezza dell'arco  $a$ ,  $\Gamma_k$  è l'insieme di archi che appartengono al percorso  $k$ ,  $L_{kl}$  è la lunghezza comune tra i percorsi  $k$  e  $l$ ,  $\delta_{al}$  è la dummy di incidenza arco-percorso, uguale a uno se il percorso  $l$  usa l'arco  $a$  e zero altrimenti, mentre  $\gamma_{CF}$  è un parametro da stimare.

I fattori di comunanza esprimono diversi concetti di similarità: l'espressione 4 dipende esclusivamente dalla lunghezza comune tra i percorsi, la 5 e 6 estendono il concetto a tutti gli archi del percorso e introducono pesi sull'importanza degli archi proporzionali al rapporto tra lunghezza dell'arco e del percorso; la 7 introduce anche i costi degli archi non condivisi e fa sì che il rapporto i fattori di comunanza di due percorsi aumentino all'aumentare della mutua sovrapposizione. I fattori di comunanza della 2 e della 5 sono sempre positivi. Conseguentemente, i parametri stimati  $\beta_{CF}$  dovrebbero essere negativi per esprimere la riduzione dell'utilità dei percorsi con archi comuni ad altri percorsi. Ancora, per percorsi unici l'argomento del logaritmo è uguale ad uno e il fattore di comunanza è nullo.

I vantaggi del modello C-Logit consistono nella robustezza desiderata nella stima dei parametri rispetto alla dimensione dell'insieme di scelta (Prato & Bekhor, 2007) e l'esistenza di una formulazione matematica equivalente per il problema SUE (Zhou & Chen, 2003). I maggiori svantaggi del C-Logit sono che il fattore di comunanza cattura solamente parte della similarità e che le regole di selezione della formula del fattore non sono suggerite. Alcuni casi di studio provano l'evidenza con alcuni risultati contro-intuitivi in termini di segno dei parametri  $\beta_{CF}$  quando i fattori di comunanza sono calcolati secondo la formule 5 e 6, mentre mostrano risultati attesi e soddisfacenti quando applicano la 7 (Ramming, 2002; Prato & Bekhor, 2007). Ancora, i valori della verosimiglianza mostrano che il Path Size Logit offre prestazioni migliori rispetto al C-Logit (Ramming, 2002; Prato & Bekhor, 2007; Prato & Bekhor, 2006).

### Path Size Logit

Ben-Akiva e Bierlaire (1999) presentano il modello Path Size Logit (PSL) per un'applicazione alla teoria di scelta discreta per le alternative aggregate, già usata in altri contesti trasportistici come la scelta della destinazione. Anche per il PSL l'espressione dell'utilità di scegliere un percorso k tra le diverse alternative riflette la semplice struttura logit:

$$P_k = \frac{\exp(V_k + \beta_{PS} \cdot \ln PS_k)}{\sum_{l \in C} \exp(V_l + \beta_{PS} \cdot \ln PS_l)} \quad (8)$$

dove  $PS_k$  e  $PS_l$  sono i path size rispettivamente dei percorsi k e l, e  $\beta_{PS}$  è un parametro da stimare. Anche se il C-Logit e il PSL hanno simili forme funzionali, ciascuno offre un'interpretazione differente in relazione al termine di correzione introdotto all'interno della funzione di utilità. Il fattore di comunanza riduce l'utilità di un percorso a causa della sua somiglianza con altri, mentre il path size indica la frazione di percorso che rappresenta

un'alternativa completa. Conseguentemente, un percorso unico ha path size uguale 1 mentre N percorsi duplicati condividono la dimensione 1/N. In letteratura sono presenti differenti formulazioni del path size (Ben-Akiva & Bierlaire, 1999; Ramming, 2002):

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al}} \quad (9)$$

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \left( \frac{L_k}{L_l} \right)^{\gamma_{PS}} \delta_{al}} \quad (10)$$

dove  $\gamma_{PS}$  è un parametro da stimare.

La formulazione originale (9) esprime il peso corrispondente alla frazione di impedenza di percorso corrispondente ad uno specifico arco come il rapporto tra le lunghezze dell'arco e del percorso (Ben-Akiva & Bierlaire, 1999). Inoltre, il termine rimanente è basato sul numero di percorsi che usano uno specifico arco e quindi è uguale a 1 per gli archi usati solo da un percorso. Si noti che l'impedenza dei percorsi che utilizzano un arco specifico non influenza questo termine, e conseguentemente la formulazione non può tenere conto del contributo di diverse dimensioni dovuto a percorsi di differente lunghezza.

La formula generalizzata (10) ha l'intento di decrementare l'influenza di percorsi eccessivamente lunghi sull'utilità di quelli brevi all'interno dell'insieme di scelta (Ramming, 2002). Inoltre, alti valori di  $\gamma_{PS}$  sembrano migliorare la bontà di adattamento del modello PSL, come dimostrato da diversi casi di studio (Ramming, 2002; Hoogendoorn-Lanser, 2005; Bekhor & Prato, 2006). Frejinger e Bierlaire (2007) dicono che la formulazione generalizzata potrebbe produrre dei risultati contro-intuitivi per elevati valori di  $\gamma_{PS}$  e quindi è da preferire la formulazione originale, anche perché è supportata da fondamenta teoriche.

La stima del PSL con la formulazione originale è un compito semplice, data la forma chiusa della struttura logit e lo sforzo computazionale limitato richiesto per il calcolo del path size. La stima del PSL con la formula generalizzata incrementa notevolmente il peso computazionale, perché la stima di  $\gamma_{PS}$  richiede la definizione di una funzione di utilità non lineare. Uno studio presenta la stima di  $\gamma_{PS}$  (Bekhor & Prato, 2006), mentre altri studi ovviano il problema computazionale ripetendo la stima del PSL per diversi valori di  $\gamma_{PS}$  e ne individuano quello ottimo (Ramming, 2002; Hoogendoorn-Lanser, 2005). Riguardo il problema computazionale, l'interpretazione comportamentale delle stime del parametro  $\gamma_{PS}$  è estremamente difficile dato che il valore stimato o ottimo varia tra 10 e 15 (Prato, 2009).

La limitazione del PSL è che il path size cattura solo parte della correlazione e rende preferibili dei modelli che tengono conto della correlazione all'interno della struttura del termine d'errore senza incrementare considerevolmente la complessità. Frejinger (2007) stima il PSL con la correzione per il campionamento considerando l'insieme totale dei percorsi, con i risultati che mostrano che stime non errate sono ottenute solo quando il termine di correzione è calcolato sull'insieme intero.

### Path Size Correction Logit

Bovy et al. (2008) rivisitano la formulazione del path size proponendo una derivazione analitica del fattore di correzione dalla teoria dell'utilità casuale basata sulle alternative aggregate e da un'approssimazione dei modelli GEV come il Paired Combinatorial Logit e il Cross Nested Logit. Il modello risultante è chiamato Path Size Correction Logit (PSCL), dove il fattore di correzione sostituisce quello della formulazione originale. Anche per il PSCL l'espressione dell'utilità di scegliere un percorso  $k$  mantiene la semplice struttura logit.

$$P_k = \frac{\exp(V_k + \beta_{PSC} \cdot \ln PSC_k)}{\sum_{l \in C} \exp(V_l + \beta_{PSC} \cdot \ln PSC_l)} \quad (11)$$

dove  $PSC_k$  e  $PSC_l$  sono rispettivamente i path size corretti dei percorsi  $k$  e  $l$ , e  $\beta_{PSC}$  è un parametro da stimare. La formula del PSC assume la seguente espressione (Bovy, et al., 2008):

$$PSC_k = - \sum_{a \in \Gamma_k} \left( \frac{L_a}{L_k} \ln \sum_{l \in C} \delta_{al} \right) \quad (12)$$

La formula pesa la lunghezza degli archi condivisi con il logaritmo del numero di percorsi che usano questi archi. La correzione path size ha un limite superiore per percorsi completamente indipendenti, esattamente come le altre formulazioni path size, mentre non ha limite inferiore perché non c'è limite superiore sul numero di percorsi che condividono un arco (Bovy, et al., 2008). Si noti che il PSC varia tra  $-\infty$  e 0, mentre quella originale varia tra 0 e 1.

La stima del PSCL è semplice come quella del PSL, perché il modello mantiene la struttura Logit originale e lo sforzo richiesto per il calcolo della correzione path size è limitata.

## 2.4.2. Strutture GEV

I modelli GEV tengono conto della similarità dei percorsi nella componente stocastica della funzione di utilità, e legano la topologia della rete a specifici coefficienti che caratterizzano la loro struttura ad albero, ma non consentono la considerazione della variazione dei gusti o la correlazione sul tempo di fattori non osservati (Prato, 2009).

### Paired Combinatorial Logit

Prashker e Bekhor (Prashker & Bekhor, 1998) adattano al contesto di scelta del percorso il modello Paired Combinatorial Logit (PCL) proposto da Chu (1989) e sviluppato poi da (Koelman & Wen, 1998). Inoltre, Prashker e Bekhor (Prashker & Bekhor, 2000) presentano una formulazione matematica per il problema SUE per la soluzione del quale è usato il PCL.

La logica alla base del modello è che i percorsi sono scelti tra una coppia di alternative all'interno dell'insieme di scelta, e la probabilità di scelta è definita come:

$$P_k = \sum_{k \neq j} P(kl)P(k|kl) \quad (13)$$

dove  $P(kl)$  è la probabilità marginale di scegliere la coppia  $(k,l)$  tra le  $n(n-1)/2$  possibili coppie, e  $P(k|kl)$  è la probabilità condizionale di scegliere il percorso  $k$  data la coppia binaria scelta  $(k,l)$ . Le probabilità condizionali e marginali dipendono dalla similarità tra i percorsi all'interno della coppia scelta:

$$P(k|kl) = \frac{\exp\left(\frac{V_k}{1-\sigma_{kl}}\right)}{\exp\left(\frac{V_k}{1-\sigma_{kl}}\right) + \exp\left(\frac{V_l}{1-\sigma_{kl}}\right)} \quad (14)$$

$$P(kl) = \frac{(1-\sigma_{kl}) \left( \exp\left(\frac{V_k}{1-\sigma_{kl}}\right) + \exp\left(\frac{V_l}{1-\sigma_{kl}}\right) \right)^{1-\sigma_{kl}}}{\sum_{p=1}^{n-1} \sum_{q=p+1}^n (1-\sigma_{pq}) \left( \exp\left(\frac{V_p}{1-\sigma_{pq}}\right) + \exp\left(\frac{V_q}{1-\sigma_{pq}}\right) \right)^{1-\sigma_{pq}}} \quad (15)$$

dove  $\sigma_{kl}$  è il coefficiente di similarità tra i percorsi  $k$  e  $l$ .

Il modello PCL introduce relazioni di similarità indipendenti per ogni coppia di alternative. Il confronto a coppie nel PCL costituisce il livello più alto della struttura a nido e quindi il numero dei nidi aumenta rapidamente con la dimensione della rete perché, teoricamente, il livello superiore include tutte le possibili coppie. La letteratura presenta due differenti

formulazioni per il coefficiente di similarità. Prashker e Bekhor (1998) definiscono un indice parametrizzato di similarità che ha alcune analogie con il fattore 4 del C-Logit:

$$\sigma_{kl} = \left( \frac{L_{kl}}{\sqrt{L_k L_l}} \right)^{\gamma_\sigma} \quad (16)$$

dove  $\gamma_\sigma$  è un parametro da stimare. Gliebe et al. (1999) introducono una differente espressione per l'indice di similarità, che dipende inoltre alla lunghezza condivisa dei percorsi della coppia:

$$\sigma_{kl} = \frac{L_{kl}}{L_k + L_l - L_{kl}} \quad (17)$$

Entrambe le equazioni limitano il coefficiente di similarità tra 0 e 1, condizione necessaria per cui il modello PCL sia consistente con la massimizzazione dell'utilità casuale. Un coefficiente uguale a uno indica che tutti gli archi del percorso sono uguali a quello di un altro, mentre un valore pari a zero indica che due percorsi non hanno nessun arco in comune. Se l'ultima condizione si ripete per ogni coppia (k,l), tutti i percorsi dell'insieme di scelta sono disgiunti e il PCL diventa un semplice MNL (Prato, 2009).

Si noti che la letteratura riporta previsioni di scelta del percorso usando il PCL per segmenti di popolazione in contesti multimodali (Benjamins, et al., 2002) e sulla valutazione della composizione dell'insieme di scelta sulle probabilità PCL (Bliemer & Bovy, 2008).

### **Cross Nested Logit**

Prashker e Bekhor (1998) adattano alla scelta del percorso il modello Cross Nested Logit (CNL) proposto da Vovsha (1997) per un'applicazione di scelta modale in un'area metropolitana vasta. Inoltre, Prashker e Bekhor (2000) forniscono una formulazione matematica equivalente al CNL per la soluzione del problema SUE.

L'assunzione alla base del modello è che i percorsi siano scelti all'interno di nidi, che fisicamente corrispondono agli archi sulla rete, e di conseguenza la probabilità di scelta è definita come:

$$P_k = \sum_m P(m)P(k|m) \quad (18)$$

dove  $P(m)$  è la probabilità marginale di scegliere il nido  $m$ , mentre  $P(k|m)$  è la probabilità condizionale di scegliere il percorso  $k$  nel nido  $m$ . Le due probabilità dipendono dai coefficienti di inclusione e di annidamento:



$$P(k|m) = \frac{(\alpha_{mk} \exp(V_k))^{1/\mu_m}}{\sum_l (\alpha_{ml} \exp(V_l))^{1/\mu_m}} \quad (19)$$

$$P(m) = \frac{\left( \sum_k (\alpha_{mk} \exp(V_k))^{1/\mu_m} \right)^{\mu_m}}{\sum_h \left( \sum_k (\alpha_{hk} \exp(V_h))^{1/\mu_m} \right)^{\mu_m}} \quad (20)$$

dove  $\alpha_{mk}$  sono i coefficienti di inclusione ( $0 \leq \alpha_{mk} \leq 1$ ) e  $\mu_m$  sono i coefficienti di annidamento ( $0 \leq \mu_m \leq 1$ ). I coefficienti di inclusione rappresentano la percentuale degli archi generici  $m$  usata dal generico percorso  $k$ , e sono soggetti a vincoli di regolarità:

$$\sum_m \alpha_{mk} = 1 \quad (21)$$

Prashker e Bekhor (1998) definiscono una relazione funzionale per l'inclusione dei coefficienti di inclusione rispetto agli archi in un percorso:

$$\alpha_{mk} = \frac{L_m}{L_k} \delta_{mk} \quad (22)$$

dove  $L_m$  è la lunghezza dell'arco (nido)  $m$ ,  $L_k$  è la lunghezza del percorso  $k$  e  $\delta_{mk}$  è la dummy di incidenza arco-percorso, uguale a uno se il percorso  $k$  usa l'arco  $m$  e zero altrimenti. Inoltre, Prashker e Bekhor (1998) assumono che tutti gli archi condividano un coefficiente di annidamento comune  $\mu$ , da stimare. Si noti che il CNL collassa nel MNL se i coefficienti  $\mu$  sono uguali a 1, mentre diventa probabilistico al livello superiore e deterministico al livello inferiore quando i coefficienti  $\mu$  sono prossimi a zero.

Esempi della stima del CNL mostrano alcuni problemi computazionali e comportamentali. Dal punto di vista del calcolo, commisurato con la sua complessità, il CNL richiede tempi di stima più lunghi rispetto ad altre specificazioni modellistiche (Ramming, 2002). Dalla prospettiva comportamentale, i coefficienti di annidamento stimati si avvicinano spesso a uno e suggeriscono che il CNL tenda a collapsare nel MNL (Ramming, 2002; Prato & Bekhor, 2006). Si noti che in questi esempi il CNL non migliora le modificazioni del MNL, e i coefficienti di annidamento prossimi a uno implicano che la similarità tra le alternative non sia catturata propriamente dalla definizione dei coefficienti di inclusione. Quindi, le performance del modello non raggiungono le aspettative desiderate (Prato, 2009).

### Generalized Nested Logit

Bekhor e Prashker (2001) adattano alla scelta del percorso la struttura del modello Generalized Nested Logit (GNL), elaborata da Wen e Koppelman (2001) come una generalizzazione della specificazione del CNL presentata da Vovsha (1997). La stessa logica del CNL si applica al GNL, e le probabilità di scelta hanno la stessa formula 18, 19 e 20. Bekhor e Prashker (2001) scrivono inoltre la formulazione matematica equivalente per il problema SUE.

Bekhor e Prashker (2001) generalizzano il modello CNL mantenendo l'espressione 22 per il coefficiente di inclusione e permettendo ad ogni nido  $m$  di avere un differente coefficienti di annidamento. La formulazione proposta assume che il coefficiente di annidamento sia un valore medio parametrizzato del coefficiente di inclusione:

$$\mu_m = \left( 1 - \frac{\sum_{l \in C_n} \alpha_{ml}}{\sum_{l \in C_n} \delta_{ml}} \right)^\gamma \quad (23)$$

dove  $\gamma$  è un parametro da stimare.

La stime dei modelli GNL mostrano similarità rispetto ai CNL. La complessità aggiuntiva dovuta alla stima di  $\gamma$  aumenta ancora lo sforzo computazionale richiesto. Ancora, i valori stimati di questo parametro sono di solito relativamente alti, tra 4 e 5 (Ramming, 2002; Prato & Bekhor, 2006): si noti che i coefficienti di annidamento tendano ad uno all'incrementare del parametro, con la conseguenza che il GNL tenda a collassare verso il MNL esattamente come il CNL in casi di studio reali (Prato, 2009).

### 2.4.3. Strutture non-GEV

Le strutture di modelli non-GEV permettono non solo modelli di sostituzione senza restrizione, ma anche la variazione casuale dei gusti e la correlazione rispetto al tempo di fattori non osservati. Siccome questi modelli non hanno un'espressione in forma chiusa per le probabilità di scelta, per la stima è necessaria la simulazione di massima verosimiglianza.

### Multinomial probit

Daganzo e Sheffi (1977) propongono il modello di scelta del percorso Multinomial Probit (MNP) assumendo una distribuzione normale per la componente casuale. La funzione di densità congiunta del termine d'errore è descritta da un vettore di medie (di lunghezza  $J$ ) e

una matrice di covarianza ( $J$ ), dove  $J$  è il numero di percorsi per una specifica coppia origine-destinazione.

La funzione cumulativa della distribuzione normale non può essere espressa in forma chiusa, quindi il calcolo delle probabilità di scelta probit di scegliere un percorso tra un elevato numero di alternative non è semplice. Il problema principale del MNP è relativo alla specificazione della matrice di covarianza, necessaria per calcolare le probabilità di scelta, che necessita di legare la varianza dei termini d'errore a parametri della rete misurabili.

Sheffi e Powell (1982) assumono che la varianza sia proporzionale a caratteristiche fissate degli archi (ad esempio la lunghezza), in modo da ottenere una distribuzione invariante del termine d'errore. Yai et al. (1997) propongono un modello MNP una matrice di covarianza strutturata per rappresentare la sovrapposizione tra percorsi alternativi. La matrice di covarianza è legata a variabili di sovrapposizione misurabili, come la lunghezza comune tra i percorsi, e il caricamento stocastico è effettuato attraverso un'integrazione numerica. Si noti che questa assunzione è essenzialmente la stessa di Sheffi e Powell (1982), solamente all'interno di un contesto differente. Infatti, Yai et al. (1997) stimano dei modelli di scelta del percorso usando matrici di covarianza strutturata con costi fissi, mentre Sheffi e Powell (1982) sviluppano la matrice di varianza nel contesto del loro programma SUE per mostrare l'equivalenza con il problema SUE (Sheffi, 1985).

L'elevato peso computazionale per la stima del MNP spiega perché la modellazione della scelta del percorso e l'assegnazione stocastica spesso adottino specificazioni alternative per rappresentare il comportamento dei viaggiatori.

### **Logit Kernel con coefficienti casuali**

La caratteristica principale del modello Logit Kernel (LK) o Mixed Logit (Ben Akiva & Bolduc, 1996; McFadden & Train, 2000) è che i fattori non osservati possono essere scomposti in una parte che contiene la correlazione e l'eteroschedasticità, mentre l'altra parte è il valore estremo identicamente e indipendentemente distribuito (i.i.d.).

La più semplice derivazione del modello LK è basata su coefficienti casuali. La probabilità per un individuo  $n$  di scegliere un percorso  $k$  ha la stessa forma del logit standard, ma è condizionale sulla distribuzione dei coefficienti  $\beta_n$  :

$$P_{nk}(\beta_n) = \frac{\exp(\beta_n' X_{nk})}{\sum_{l \in C_d} \exp(\beta_n' X_{nl})} \quad (24)$$

dove  $X_{nk}$  e  $X_{nl}$  sono variabili osservate che si riferiscono all'individuo  $n$  e rispettivamente alle alternative  $k$  e  $l$ , mentre  $\beta_n$  è un vettore di coefficienti casuali rappresentante i gusti dell'individuo  $n$ . Siccome il modellista non conosce  $\beta_n$  e non può imporre la condizione su  $\beta_n$ , la probabilità incondizionata di scegliere il percorso  $k$  è l'integrale di  $P_{nk}(\beta_n)$  su tutti i possibili valori di  $\beta_n$ :

$$P_{nk} = \int \frac{\exp(\beta' X_{nk})}{\sum_{l \in C_d} \exp(\beta' X_{nl})} f(\beta) d\beta \quad (25)$$

dove  $f(\beta)$  è la densità della distribuzione di  $\beta$  sulla popolazione. La probabilità incondizionata è calcolata attraverso simulazione:

$$P_{nk} = \frac{1}{D} \sum_{d=1}^D \frac{\exp(\beta'_d X_{nk})}{\sum_{l \in C_d} \exp(\beta'_d X_{nl})} \quad (26)$$

dove  $\beta'_d$  indica un'estrazione  $d$  dalla distribuzione di  $\beta'$  e  $D$  è il numero di estrazioni.

Nonostante la corrente discussione (dei ricercatori sulla scelta del percorso e sui metodi di assegnazione) sull'adeguatezza delle distribuzioni dei coefficienti, solo pochi studi mostrano stime di modelli di scelta del percorso tenendo conto dell'eterogeneità dei gusti e dell'eteroschedasticità. Ben Akiva et al. (1993) stimano un modello con due alternative e dati cross-sectional dove il coefficiente associato al tempo è distribuito secondo una log-normale e l'integrale è valutato con una quadratura Gaussiana. Dial (1997) considera, all'interno di un algoritmo per l'assegnazione del traffico, che i guidatori abbiano diverse percezioni sui tempi di viaggio e sui costi a causa dell'abitudine, gusti o informazioni. Nielsen (2000) dice che le distribuzioni log-normali e gamma sono adatte per simulare le preferenze in caso di assegnazione stocastica del traffico con differenze nelle funzioni di utilità dei passeggeri. Han et al. (2001) usano distribuzioni uniformi e normali per il ritardo e il tempo di viaggio su una rete con un elevato livello di traffico per modellare preferenze dichiarate (SP) di scelte di percorsi a coppie, e aggiungono che la distribuzione log-normale non produce risultati soddisfacenti. Jou (2001) investiga sull'impatto di informazioni pre-spostamento sul comportamento di scelta del percorso attraverso la stima di un modello con componente casuale con il tempo di viaggio distribuito secondo una normale. Lam e Small (2001) modellano la scelta tra un percorso libero e uno a pagamento, tenendo conto della mediana del tempo di viaggio e della variabilità dello stesso in relazione al periodo del giorno. Nielsen et al. (2002) stimano un modello per differenti categorie di guidatori testando coefficienti

per il tempo di viaggio e costo distribuiti secondo una normale e log-normale. Nielsen (2004) valuta l'eterogeneità nelle risposte dei guidatori rispetto a schemi di tariffazione attraverso un esperimento SP sul road pricing e modellando le preferenze rilevate acquisite tramite dispositivi GPS.

Dato che ci si aspetta che i segni del tempo di viaggio e dei costi siano negativi, la distribuzione log-normale dovrebbe essere da preferire rispetto alla illimitata distribuzione normale che implica che alcuni utenti possano preferire percorsi più lunghi e costosi. I risultati dei casi di studio mostrano che i modelli con i coefficienti log-normali o non producano risultati soddisfacenti (Han, et al., 2001) oppure introducano eccessiva variabilità, producendo talvolta delle scelte illogiche (Nielsen, et al., 2002), mentre i coefficienti distribuiti secondo la gamma sono da preferire perché riproductivi e non negativi (Nielsen, 2000). Differenti applicazioni di modelli di scelta discreta introducono delle distribuzioni limitate per i coefficienti di variabili con segno atteso (vedere ad esempio (Train & Sonnier, 2005; Hess, et al., 2005).

Nella scelta del percorso, i modelli con coefficienti di variazione del costo sono attualmente diversi dai modelli con coefficienti del tempo variabili, perché la varianza complessiva cambia con il costo nel primo caso e con il tempo nell'altro, causando che le assunzioni sull'eteroschedasticità siano differenti (Nielsen, et al., 2002). Caussade et al. (2005) modellano l'eteroschedasticità nella scelta del percorso mentre esamina l'influenza della complessità di schemi di scelte dichiarate sull'abilità di scegliere del consumatore.

### **Logit Kernel con un approccio ai fattori analitici**

Un modello LK può essere usato senza un'interpretazione a coefficienti casuali, semplicemente rappresentando componenti d'errore che creano correlazione tra le utilità per le differenti alternative.

Il primo adattamento del modello LK alla scelta del percorso assume che la covarianza dell'utilità dei percorsi sia proporzionale alla lunghezza per la quale i percorsi si sovrappongono (Bekhor, et al., 2002). Costruendo sulla derivazione del LK con approccio ai fattori analitici, la probabilità di scegliere il percorso  $k$  dato un vettore  $\zeta$  di variabili normali standard è la seguente:

$$P_k = \Lambda(k | \zeta) = \frac{\exp(\mu(X_k \beta + F_k T \zeta))}{\sum_{l \in C_n} \exp(\mu(X_l \beta + F_l T \zeta))} \quad (27)$$

dove  $\beta_{(1 \times B)}$  è il vettore colonna dei parametri,  $X_k$  è la k-esima riga della matrice delle variabili esplicative  $X_{(j \times B)}$ ,  $F_k$  è la k-esima riga della matrice dei fattori di caricamento  $F_{(j \times M)}$  (J percorsi e M elementi della rete),  $T_{(M \times M)}$  è una matrice diagonale di parametri di covarianza  $\sigma_m$ ,  $\zeta_{(M \times 1)}$  è un vettore di variabili normali standard. Bekhor et al. (2002) assumono che i fattori specifici dell'arco sono normali i.i.d., la varianza è proporzionale alla lunghezza dell'arco, la matrice F corrisponde alla matrice di incidenza arco-percorso e la matrice T corrisponde alla matrice di varianza arco-fattore. In accordo, viene stimato il parametro di covarianza  $\sigma$  condiviso da ogni arco.

Il secondo adattamento del modello LK alla scelta del percorso con componenti d'errore propone una prospettiva differente e interessante (Frejinger & Bierlaire, 2007). Il modello assume che due percorsi che passano attraverso la stessa porzione di rete potrebbero condividere alcuni attributi ed essere simili, anche senza condividere nessun arco. Frejinger e Bierlaire (2007) definiscono la matrice F come la matrice di incidenza percorso-porzione di rete e T come matrice di covarianza associata ai componenti della porzione di rete. Di conseguenza, sono stimati dei parametri unici di covarianza  $\sigma_m$  per ogni componente della porzione di rete. I modelli LK sono impegnativi in termini di tempi computazionali, perché il vettore  $\zeta$  è sconosciuto le probabilità non condizionate sono calcolate attraverso la simulazione:

$$P_k = \int_{\zeta} \Lambda(k | \zeta) \prod_{m=1}^{M_n} \varphi(\zeta_n) d\zeta_n = \frac{1}{D} \sum_{d=1}^D \Lambda(k | \zeta^d) \quad (28)$$

dove  $\varphi(\zeta_n)$  è la funzione di densità normale standard,  $\zeta^d$  indica un'estrazione d dalla distribuzione di  $\zeta$  e D è il numero di estrazioni.

Ci sono tuttavia per entrambi altri problemi per l'adattamento del modello LK con l'approccio a fattori analitici. Quando è stimato il parametro di covarianza condiviso da ogni arco nell'insieme di scelta, Ramming (2002) dimostra l'instabilità delle stime anche per un numero elevato di estrazioni. Quando sono stimati i parametri di covarianza di ogni componente della sotto-rete, le stime della covarianza e le prestazioni del modello appaiono molto dipendenti dalla discrezione dell'analista nella selezione di queste componenti. Non sono fornite regole sistematiche per la definizione dei componenti della sotto-rete ed è necessaria la stima di modelli di test per individuare la struttura più significativa e appropriata delle matrici F e T. Di conseguenza, il costo computazionale incrementa notevolmente a causa della stima del modello test, anche per bassi valori di D estrazioni casuali per la simulazione dell'integrale.

### **3. DESCRIZIONE DEI DATI**

#### **3.1. Introduzione**

Di seguito è riportata una sintesi delle attività e dei dati raccolti durante la campagna d'indagine "Casteddu Mobility Styles" (CMS nel seguito), svolta tra il 2011 e il 2012 in un ambito territoriale comprendente la città di Cagliari (Italia) e la sua area vasta. Da questi dati infatti possono essere acquisite anche tutte le informazioni relative alla descrizione spaziale e temporale dei percorsi scelti e realmente eseguiti.

Il fine principale di CMS è quello di analizzare il comportamento di viaggio degli utenti del sistema di trasporto:

- per identificare quali siano i fattori in grado di generare, nei confronti di utenti utilizzatori dell'automobile, dei cambiamenti volontari di comportamento di viaggio, in particolare per ciò che riguarda la scelta del modo di trasporto con il quale effettuare i propri spostamenti quotidiani;
- per valutare l'efficacia di una cosiddetta "soft measure" (in particolare un programma personalizzato di viaggio)

Nell'area di studio presa in esame, infatti, è operativa una linea di metropolitana leggera (in esercizio dal 2008, denominata "Metrocagliari") che, allo stato attuale, opera per circa il 25% della propria capacità, in un corridoio dove quotidianamente si registrano circa 150.000 spostamenti in automobile.

Nello specifico, quindi, l'obiettivo si traduce nell'implementazione di una misura che cerca di favorire l'utilizzo della metropolitana, disincentivando, allo stesso tempo, l'auto privata.

Per comprendere quali fossero i fattori (leve) principali sui quali concentrare l'attenzione per sensibilizzare gli individui a cambiare volontariamente il proprio comportamento di viaggio, sono state analizzate le caratteristiche degli utenti che già usavano il sistema di metropolitana, al fine di identificare quali fossero quegli aspetti che, in un contesto in cui l'automobile privata è il mezzo di trasporto prevalente, li avessero spinti a cambiare modalità. Questo campione di utenti è stato selezionato attraverso la somministrazione di un questionario online, dettagliatamente descritto nel seguito del documento.

Successivamente è stato selezionato un secondo campione di individui aventi le stesse caratteristiche dei precedenti (socio-economiche, di residenza, zona della città dove svolgere l'attività di lavoro/studio), con l'unica differenza di essere utilizzatori dell'auto, quindi gli utenti ideali a cui sottoporre i programmi di informazione e sensibilizzazione al cambiamento comportamentale di viaggio.

L'indagine è stata strutturata in due fasi, della durata di una settimana ciascuna: nella prima, ogni utente del secondo campione selezionato è stato monitorato (24/24 h) attraverso un sistema denominato Activity Locator (nel seguito descritto nel dettaglio), in grado di registrare, attraverso un'applicazione implementata in uno smartphone dotato di GPS, tutte le attività e gli spostamenti da esso effettuati.

Al termine dei primi sette giorni di monitoraggio, i dati raccolti sono stati analizzati per comprendere quali fossero i comportamenti di attività e viaggio dell'utente auto, al fine di predisporre un programma personalizzato di viaggio nel quale, facendo leva su quegli aspetti identificati come primari per il cambiamento comportamentale, si presentavano delle alternative alle scelte effettuate, mostrandone la convenienza e l'efficacia attraverso confronti e dati oggettivi. L'individuo è stato quindi monitorato per una seconda settimana, durante la quale si è avuta la possibilità di verificare se ed in che misura le proposte presentate nel programma personalizzato avessero avuto effetto (cambio comportamentale da modo auto a modo auto+metro).

Hanno partecipato all'indagine un totale di 109 utenti, monitorati 24/24 h per un periodo complessivo di 14 giorni, durante i quali, grazie al sistema GPS implementato negli smartphone, è stato possibile registrare tutti i percorsi effettuati e tutte le caratteristiche ad essi associate (tempi, distanze, modalità, motivo dello spostamento, etc.), per un totale di 8831 spostamenti.

### **3.2. Casteddu Mobility Styles**

L'indagine ha riguardato l'ambito territoriale di Cagliari (Italia) e l'area vasta circostante, con una partecipazione di un campione di 109 utenti, selezionati secondo le modalità illustrate nel seguito. Di questi, 23 (reference group) erano già utilizzatori della metropolitana, in particolare nella modalità Park and Ride, mentre i restanti 86 erano utenti che usavano l'auto privata, le cui caratteristiche di spostamento, residenza e lavoro sono risultate coerenti con la possibilità di utilizzare la modalità P&R, rendendoli così i veri destinatari della ricerca.

CMS è stato strutturato in tre fasi:

- 1) una prima indagine campionaria destinata agli attuali utilizzatori della metro e intitolata "Chi usa Metrocagliari?" (basata sul "Questionario 0" – allegato);
- 2) una seconda indagine campionaria denominata "Abitudini di viaggio" (basata sul "Questionario 1" – allegato) volta ad intercettare un campione di auto-guidatori



identificabili come potenziali utilizzatori della Metro da sottoporre al programma di cambiamento volontario di viaggio;

- 3) una campagna di rilievi con l'Activity Locator.

### **3.2.1. Selezione 1 - Chi usa Metrocagliari?**

L'obiettivo di questa prima indagine campionaria, avvenuta da Febbraio ad Aprile 2011, era principalmente quello di conoscere quali fossero gli elementi caratteristici accomunanti gli utilizzatori abituali della metro, in modo da avere dei riferimenti da analizzare per strutturare le proposte di cambiamento da sottoporre agli utenti della modalità "car-only". A questo scopo, gli attuali utilizzatori sono stati "intercettati" direttamente a bordo della metropolitana, ai quali è stata poi consegnata una cartolina informativa con l'invito, qualora fossero stati interessati, alla compilazione di un questionario online (Questionario 0). Sono state quindi acquisite:

- le caratteristiche socio-economiche, demografiche, comportamentali, attitudinali dell'attuale utenza Metrocagliari;
- le modalità di utilizzo della Metro da parte dei vari segmenti identificabili nel campione, i costi e benefici dichiarati, le barriere percepite all'utilizzo della metro, la concorrenzialità con l'utilizzo dell'auto privata.

Il Questionario 0, pubblicato online sul sito del programma, era suddiviso in 6 sezioni:

- 1) descrizione generale, tra cui alternative disponibili, motivo dello spostamento (che stavano effettuando quando hanno ricevuto la cartolina), frequenza dello spostamento, etc,;
- 2) descrizione dello spostamento;
- 3) cambiamento comportamentale (circostanze che hanno spinto l'utente ad utilizzare la metropolitana);
- 4) attitudine pro ambientale e al movimento;
- 5) informazioni personali e familiari;
- 6) informazioni sull'utilizzo dell'automobile.

Dall'analisi delle risposte ai quesiti, si è rilevato come una quota pari al 17% del campione intervistato utilizzasse la Metro in modalità Park & Ride (P&R), recandosi quindi alla fermata di salita con la propria auto e raggiungendo la propria destinazione in metropolitana (di solito verso la zona circostante il capolinea della linea (fermata "Repubblica")). L'analisi delle

caratteristiche e delle risposte relative a questo particolare sub-campione di metro-users ha permesso allora di identificare un determinato target di comportamento e di stile di mobilità sostenibile, quello dei Park and Riders (P&Rs), che si è deciso di proporre agli automobilisti che, a dispetto di caratteristiche molto simili, non avessero ancora individuato la stessa alternativa sostenibile (potenziali Park and Riders, PP&Rs). La scelta di utilizzare le caratteristiche di questo target permette inoltre di offrire un'alternativa che non comporti una rinuncia totale dell'utilizzo della propria auto, che come dimostrato da numerosi studi, costituisce un simbolo affettivo fortissimo per alcuni individui. Del sottocampione selezionato hanno partecipato alla fase finale di CMS 23 utenti P&R.

### **3.2.2. Selezione 2 - Abitudini di viaggio**

L'obiettivo della seconda indagine campionaria era quello di intercettare un segmento di utenti utilizzatori dell'automobile che avessero le caratteristiche per essere considerati potenziali utilizzatori della metro in modalità Park and Ride (PP&Rs). Attraverso diverse forme di pubblicità per l'iniziativa, gli utenti interessati potevano compilare un questionario online (Questionario 1), dall'analisi del quale sarebbero stati poi scelti i possibili partecipanti alla fase finale.

Il questionario è stato strutturato in tre sezioni relative a:

- area di residenza;
- possesso di patente e possesso d'auto;
- frequenza e motivo dello spostamento verso la zona del capolinea della metro.

Tali domande hanno permesso di identificare quali fossero gli automobilisti con caratteristiche simili agli attuali Park and Riders, che per origine e destinazione dei propri spostamenti avrebbero potuto integrare l'utilizzo della metro all'interno del proprio stile di mobilità.

Al termine dell'analisi delle risposte fornite, tra gli utenti aventi le caratteristiche più simili al target di riferimento, 86 PP&Rs hanno partecipato alla fase finale di CMS.

Ultimate le due fasi di selezione sono stati quindi individuati 109 partecipanti alla fase finale di CMS, che prevedeva il monitoraggio, per due settimane, tramite l'AL.

Tutti gli utenti sono stati invitati alla compilazione del "Questionario 2" (in allegato), lo scopo del quale era quello di integrare le informazioni già in possesso, con la richiesta di informazioni generali su:

- luogo di residenza;
- impegno nella riduzione delle emissioni inquinanti;
- attitudine nei confronti dell'ambiente;
- caratteristiche personali;
- possesso dell'auto;
- caratteristiche dei propri spostamenti verso la fermata "Repubblica" (capolinea della linea di metropolitana).

In particolare, ai PP&Rs sono state richieste anche tutte le informazioni socio-economiche e di attività che per i P&R erano già state acquisite attraverso il Questionario 0.

In definitiva, per tutti i 109 utenti si hanno a disposizione le seguenti informazioni:

- socio-economiche-personali (età, sesso, occupazione, livello di istruzione, reddito, stato civile, eventuali figli, possesso di patente, possesso dell'automobile);
- residenza;
- spostamento effettuato più frequentemente;
- abitudini di viaggio (mezzi utilizzati, orari, luogo di parcheggio);
- attitudine pro ambientale (impegno quotidiano per la riduzione delle emissioni, spesa media per gli spostamenti, possesso e utilizzo della bicicletta, pratica di sport, propensione agli spostamenti a piedi, conoscenza del servizio di metropolitana),
- informazioni sull'eventuale veicolo in possesso (modello, anno di immatricolazione, anno d'acquisto, km percorsi all'anno, conoscenza della quantità delle emissioni).

### **3.3. Campagna di rilievo**

Come illustrato in precedenza, il lavoro si è sviluppato in due settimane: nella prima (week before) sono stati raccolti i dati spaziali e temporali (attraverso l'AL) sugli attuali comportamenti di viaggio degli utenti e sulle attività da essi compiute durante la giornata, mentre nella seconda settimana (week after) sono stati rilevati gli stessi dati per la verifica dei comportamenti conseguenti alla somministrazione di un PTP (Personalised Travel Plan), avvenuta tra la fine della prima settimana e l'inizio della seconda.

### **3.3.1. Week “before”**

Prima di iniziare la fase di monitoraggio, ad ogni partecipante è stato assegnato un tutor (il quale poteva controllare contemporaneamente un massimo di 5 utenti), che ha provveduto alla somministrazione di un questionario di “consegna” (Questionario 3 – in allegato), nel quale era richiesto di motivare alcune risposte date alle domande del Questionario 2 tra cui:

- l’attenzione all’ambiente;
- l’uso dell’automobile;
- le caratteristiche dello spostamento verso il capolinea della metro (fermata Repubblica);
- i fattori che secondo lui potrebbero incentivare l’utilizzo della metro;
- le caratteristiche dei propri spostamenti (orario partenza/arrivo, orario inizio/fine lavoro, caratteristiche dell’itinerario utilizzato).

Al termine della prima settimana di monitoraggio con l’AL, a ciascun partecipante è stato proposto un piano personalizzato di viaggio (PTP) attraverso il quale lo si incoraggiava ad integrare la metro nei suoi schemi di attività e di viaggio giornalieri, riducendo così gli spostamenti svolti con l’auto privata, ed eventualmente ad operare uno switch modale da auto a Park and Ride. In tal modo si assume che gli individui, a seguito della somministrazione del PTP, possano riconsiderare le proprie scelte di viaggio.

### **3.3.2. Week “after”**

Nella seconda settimana di monitoraggio gli utenti proseguono con l’utilizzo dell’Activity Locator. Al termine, gli è stato somministrato il questionario di fine indagine (Questionario 4 – in allegato), nel quale sono state richieste le impressioni riguardo alla partecipazione al programma, alla sua validità, all’impegno richiesto rispetto all’utilizzo dello strumento.

## **3.4. Activity Locator**

Per il monitoraggio degli utenti partecipanti alla campagna d’indagine è stato utilizzato l’Activity Locator (AL) (Meloni, et al., 2011), strumento sviluppato e testato dal CIREM (Centro Ricerche Economiche e Mobilità dell’Università di Cagliari) nel corso del 2009.

L’Activity Locator è un efficiente ed economico sistema di tracciamento e storicizzazione dei percorsi quotidiani degli individui attraverso un GPS integrato su cellulari smartphone di ultima generazione (piattaforme Symbian e/o Android). Tecnicamente, l’architettura

dell'Activity Locator è costituita da due parti: un software client e un software server. Su ogni cellulare è installata un'applicazione software "Client" (in quanto installabile su diversi dispositivi) sviluppata in linguaggio C++/Java che costituisce l'interfaccia attraverso cui ogni utente dichiara (attraverso una sequenza di menù a tendina) la sequenza di attività e spostamenti svolti durante il giorno con i relativi attributi, esattamente come in un classico diario di attività, ma con la possibilità di farlo in tempo reale e non la sera al rientro a casa.

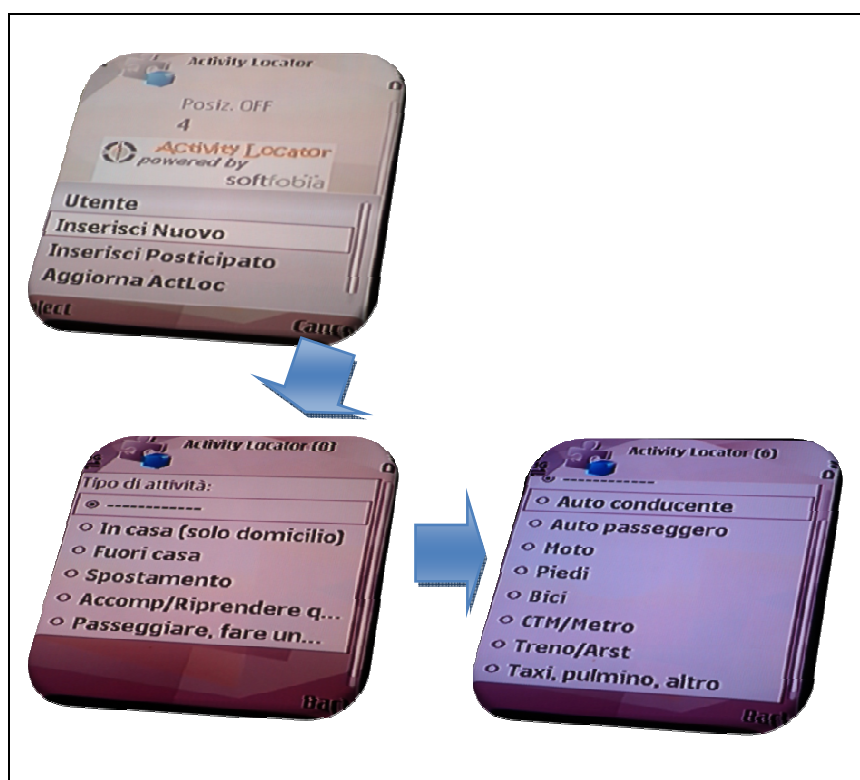


Figura 4 - Applicazione software "Client" - Inserimento dati

Le informazioni raccolte sono di due tipi: (1) informazioni georeferenziate spazio-temporali raccolte tramite il dispositivo GPS che rappresentano tutti i percorsi effettuati dagli individui nel periodo di indagine, comprese quindi localizzazioni dell'origine degli spostamenti, della destinazione, degli stop intermedi effettuati, del mezzo utilizzato e della compagnia con cui ogni spostamento è stato effettuato, (2) informazioni delle attività svolte che includono sia attività in casa che fuori casa, tipologia e compagnia con cui sono state. Nella figura seguente è illustrato lo schema strutturale dell'applicazione Client.

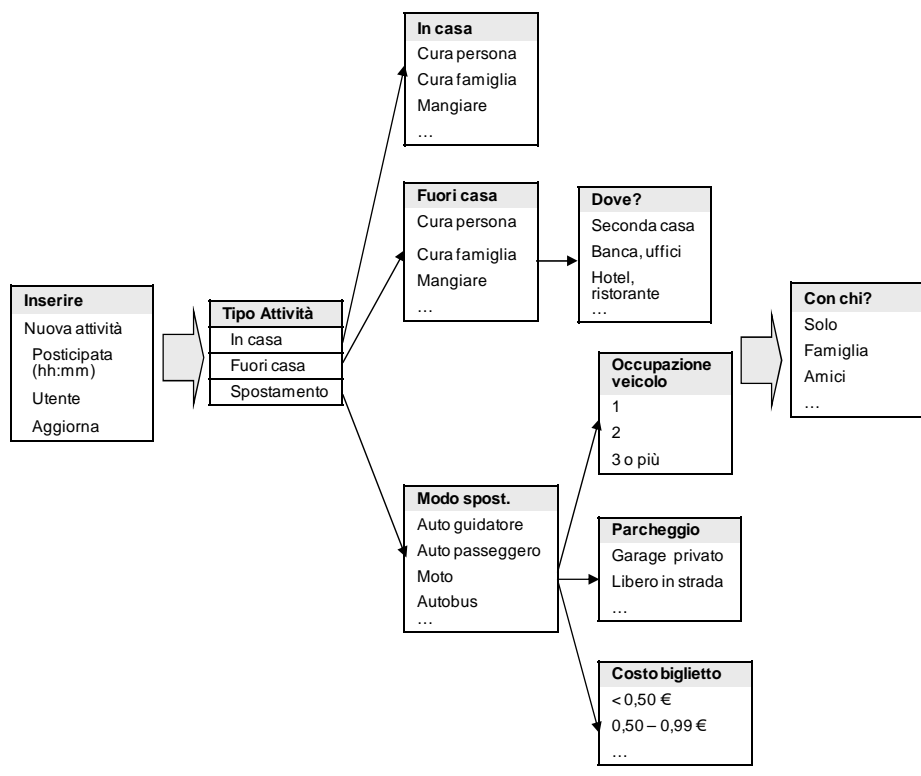


Figura 5 - Struttura Applicazione Client

L'applicazione costituisce l'interfaccia tra l'utente e un software "Server" che raccoglie e storicizza le informazioni inviate attraverso una semplice ed economica connessione Internet del cellulare.

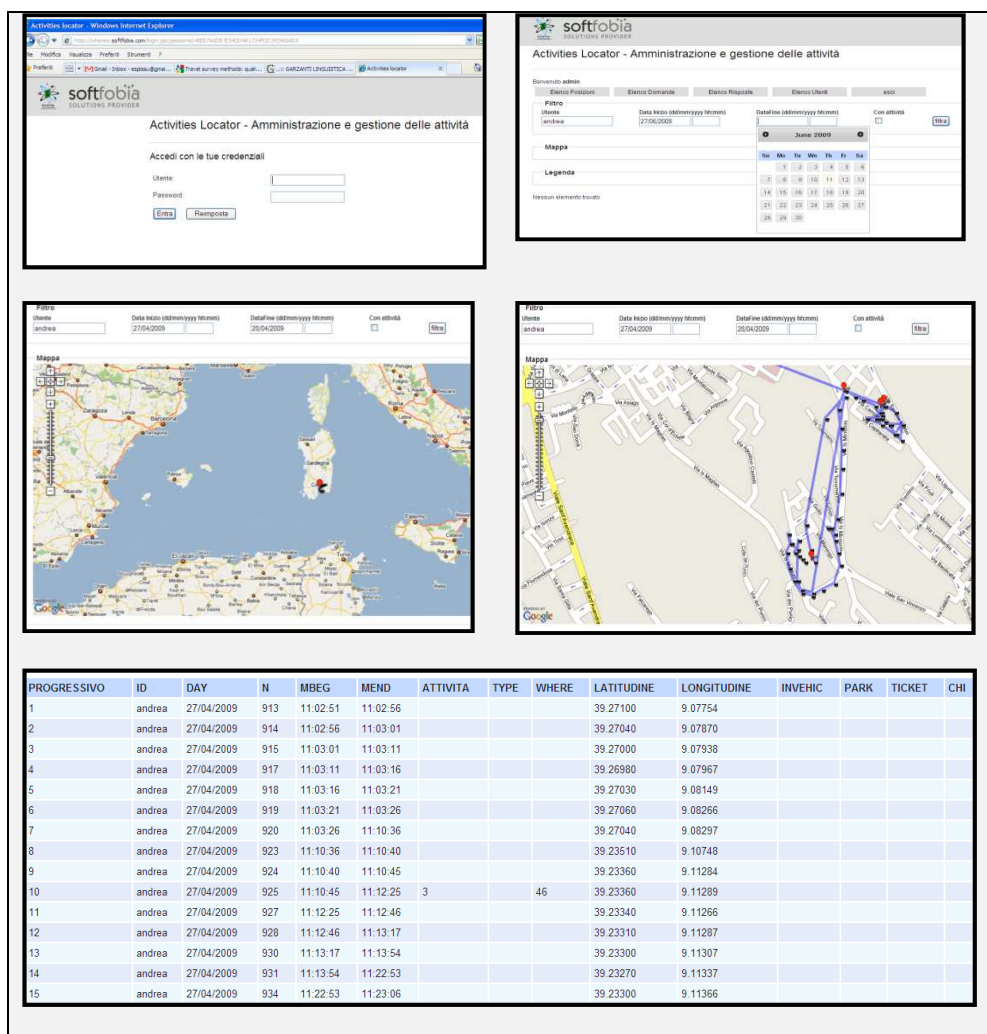


Figura 6 - Struttura software "Server"

Ogni utente, e così pure gli amministratori dell'indagine, può accedere facilmente alle informazioni raccolte attraverso l'autenticazione (nome utente, password). Ogni partecipante è individuabile in tempo reale, per mezzo di un simbolo (localizzato su una mappa) che ha ad esso associate le informazioni inserite manualmente dall'utente (attività e spostamenti) e quelle di posizionamento (ad opera del dispositivo GPS). Tutte le informazioni raccolte vengono organizzate sottoforma di database che può essere agevolmente scaricato su un qualsiasi computer in formato .xls. I dati di posizionamento, una volta sovrapposti alla rete urbana con l'ausilio di un qualsiasi software GIS, restituiscono i percorsi quotidiani di tutti gli utenti. In questa ricerca, l'Activity Locator è stato installato su smartphone Nokia N95, dotati di un sistema GPS integrato. Recentemente, gli sviluppatori hanno preparato una versione del software applicativo compatibile con il sistema operativo iOS, di proprietà della Apple, in modo da avere a disposizione anche un'applicazione per iPhone, ormai diffusissimo sul mercato.

### 3.5. Dati raccolti

Per tutti i 109 utenti, monitorati per 14 giorni ciascuno, 24/24h, è stato possibile raccogliere informazioni di due tipi: attività e spostamenti. Per le prime sono disponibili il luogo di svolgimento, l'orario di inizio e di fine (e quindi la durata), la compagnia e il tipo di attività stessa; mentre, per ciò che riguarda gli spostamenti (ne sono stati registrati 8831 complessivi), grazie al dispositivo GPS è stato possibile ottenere, oltreché il tracciato spaziale, anche le informazioni temporali ad esso legate come l'orario di inizio e fine.

Inoltre, per mezzo dell'applicazione software, è stato possibile ottenere le informazioni riguardanti il modo di trasporto, la tipologia di documento di viaggio (biglietto/abbonamento e prezzo indicativo), il numero di persone a bordo dell'auto, il tipo di parcheggio (libero o a pagamento, con la possibilità di indicare la tariffa).

Ogni tracciato GPS (costituito da una sequenza di punti "posizione" referenziati) è stato poi sottoposto a delle tecniche di map-matching, attraverso le quali è stato possibile associare ogni "punto GPS" ad un arco della rete, creando così degli shape-file di linee (uno per ogni utente), dove ad ogni percorso corrisponde una linea alla quale sono associati tutti gli attributi ad esso relativi.

Di seguito sono riportate due tabelle di sintesi, una per le attività e una per gli spostamenti, le quali mostrano la tipologia di informazioni a disposizione nell'Activity Locator.



<p><b>Tipo di attività</b></p> <ul style="list-style-type: none"> <li>• in casa (solo domicilio)</li> <li>• fuori casa</li> <li>• spostamento</li> </ul> <p><b>Luogo di svolgimento</b></p> <ul style="list-style-type: none"> <li>• seconda casa, da amici</li> <li>• centro estetico, parrucchiere.</li> <li>• hotel, ristorante, bar, pub</li> <li>• medico, ospedale</li> <li>• circolo, associazione, chiesa</li> <li>• fermata bus, stazione</li> <li>• studio medico</li> <li>• seconda casa</li> <li>• posto di lavoro/studio</li> <li>• negozio, centro commerciale</li> <li>• uffici, distributore, meccanico</li> <li>• evento, cinema, stadio</li> <li>• palestra, campo sportivo</li> <li>• aperto (parco, spiaggia)</li> </ul>	<p><b>Compagnia</b></p> <ul style="list-style-type: none"> <li>• da solo/a</li> <li>• con familiari</li> <li>• con amici, colleghi</li> <li>• con animale domestico</li> <li>• con familiari e amici/colleghi</li> </ul> <p><b>Attività</b></p> <ul style="list-style-type: none"> <li>• prepararsi, lavarsi</li> <li>• cura famiglia, casa</li> <li>• mangiare, preparazione pasti</li> <li>• lavoro/studio</li> <li>• acquisti (internet, pay-tv)</li> <li>• conto online, pagamenti</li> <li>• attività ricreativa, socializzare.</li> <li>• sport, attività fisica</li> <li>• svago (tv, libri, internet)</li> <li>• acquisti, fare la spesa</li> <li>• commissioni (auto, casa)</li> <li>• volontariato, religione</li> <li>• attesa (passaggio o altro)</li> <li>• dormire, relax, riposo</li> <li>• pausa caffè, bibita</li> <li>• salita/discesa passeggeri</li> <li>• passeggiare</li> </ul>
--	---

Tabella 3 - Informazioni sulle attività

<p><b>Modalità</b></p> <ul style="list-style-type: none"> <li>• auto conducente</li> <li>• auto passeggero</li> <li>• moto</li> <li>• piedi</li> <li>• bici</li> <li>• ctm</li> <li>• treno/arst</li> <li>• taxi, pulmino, altro</li> <li>• pulmino/altro</li> </ul> <p><b>Persone a bordo</b></p> <ul style="list-style-type: none"> <li>• 1</li> <li>• 2</li> <li>• 3</li> <li>• 4 o Più</li> </ul> <p><b>Biglietto</b></p> <ul style="list-style-type: none"> <li>• biglietto metro – treno</li> <li>• comprato a bordo</li> <li>• corsa singola 90 min</li> <li>• carnet 12 corse</li> <li>• biglietto a tempo</li> <li>• biglietto giornaliero</li> <li>• carta settimanale</li> <li>• abbonamento mensile o annuale</li> </ul>	<p><b>Parcheggio</b></p> <ul style="list-style-type: none"> <li>• parcheggio privato, garage</li> <li>• lungo strada libero</li> <li>• lungo strada pagamento</li> <li>• area sosta libera</li> <li>• area sosta a pagamento</li> </ul> <p><b>Prezzo Biglietto/Parcheggio</b></p> <ul style="list-style-type: none"> <li>• gratis</li> <li>• meno di 50 cent</li> <li>• 50 cent – 0.99 euro</li> <li>• 1 – 1.99 euro</li> <li>• 2 – 2.99 euro</li> <li>• 3 – 4.99 euro</li> <li>• 5 euro o più</li> <li>• 5 - 9.99 euro</li> <li>• 10 - 14.99 euro</li> <li>• 15 - 19.99 euro</li> <li>• 20 - 29.99 euro</li> <li>• 30 euro e più</li> </ul>
--	--

Tabella 4 - Informazioni sugli spostamenti

<p><b>Residenza</b></p> <ul style="list-style-type: none"> <li>• comune</li> <li>• indirizzo</li> </ul> <p><b>Caratteristiche socio-economiche</b></p> <ul style="list-style-type: none"> <li>• età</li> <li>• sesso</li> <li>• occupazione</li> <li>• livello di istruzione</li> <li>• reddito</li> <li>• stato civile</li> <li>• eventuali figli</li> <li>• possesso di patente</li> <li>• possesso dell'automobile</li> </ul> <p><b>Informazioni sull'automobile in possesso</b></p> <ul style="list-style-type: none"> <li>• modello</li> <li>• anno di produzione</li> <li>• anno d'acquisto</li> <li>• km percorsi annualmente</li> <li>• grammi di CO2 emessi</li> <li>• premio assicurazione</li> </ul> <p><b>Spostamenti verso Piazza Repubblica</b></p> <ul style="list-style-type: none"> <li>• frequenza</li> <li>• motivazione</li> <li>• mezzo utilizzato in alternativa all'auto</li> <li>• attività svolta a destinazione</li> <li>• luogo e tipo di parcheggio (in caso di modo auto)</li> </ul> <p><b>Incidenza dei fattori che hanno influenzato la scelta dell'auto</b></p> <ul style="list-style-type: none"> <li>• minor tempo di viaggio</li> <li>• possibilità di scegliere l'ora in cui uscire di casa e da lavoro</li> <li>• accompagnare/riprendere qualcuno, fare altre attività etc.</li> <li>• scomodità del trasporto pubblico</li> <li>• assenza di alternative</li> </ul>	<p><b>Impegno per la riduzione delle emissioni da*</b></p> <ul style="list-style-type: none"> <li>• trasporti</li> <li>• energia</li> <li>• rifiuti</li> <li>• abitudini quotidiane</li> <li>• tecnologia</li> </ul> <p>*(nulla, poco, moderatamente, molto, estremamente, non so)</p> <p><b>Utilizzo del trasporto pubblico in famiglia*</b></p> <p>*(nessuno, coniuge, figli, genitori, fratelli, sorelle)</p> <p><b>Raggiungimento dell'intorno di Piazza Repubblica</b></p> <ul style="list-style-type: none"> <li>• park and ride</li> <li>• automobile</li> </ul> <p><b>Attitudine pro-ambientale</b></p> <ul style="list-style-type: none"> <li>• spesa mensile per gli spostamenti*</li> </ul> <p>*(meno di 50 €, 50 - 100 €, 100 - 300 €, piu' di 300 €)</p> <ul style="list-style-type: none"> <li>• pratica di sport settimanale</li> <li>• possesso di bicicletta e frequenza di utilizzo</li> <li>• minuti di camminata giornalieri</li> </ul> <p><b>Conoscenza del servizio Metrocagliari</b></p> <ul style="list-style-type: none"> <li>• indicazione della fermata più vicina alla propria abitazione</li> <li>• presa in considerazione del P&amp;R</li> <li>• motivo di non utilizzo del P&amp;R</li> </ul> <p><b>Fattori che incentiverebbero l'utilizzo di mezzi alternativi all'auto</b></p> <ul style="list-style-type: none"> <li>• tasso di ingresso alla zona</li> <li>• assenza di posteggi per l'auto</li> <li>• possibilità di park and ride</li> </ul> <p><b>Spostamento più frequente</b></p> <ul style="list-style-type: none"> <li>• origine e destinazione</li> <li>• orario partenza</li> <li>• orario arrivo</li> <li>• orario inizio e fine lavoro</li> <li>• arrivo in anticipo al lavoro (ed eventuale attività svolta nell'attesa)</li> <li>• arterie principali percorse</li> <li>• motivo della scelta di quel percorso</li> <li>• eventuali spostamenti frequenti verso la Piazza Repubblica</li> </ul>
--	---

Tabella 5 - Dati questionari

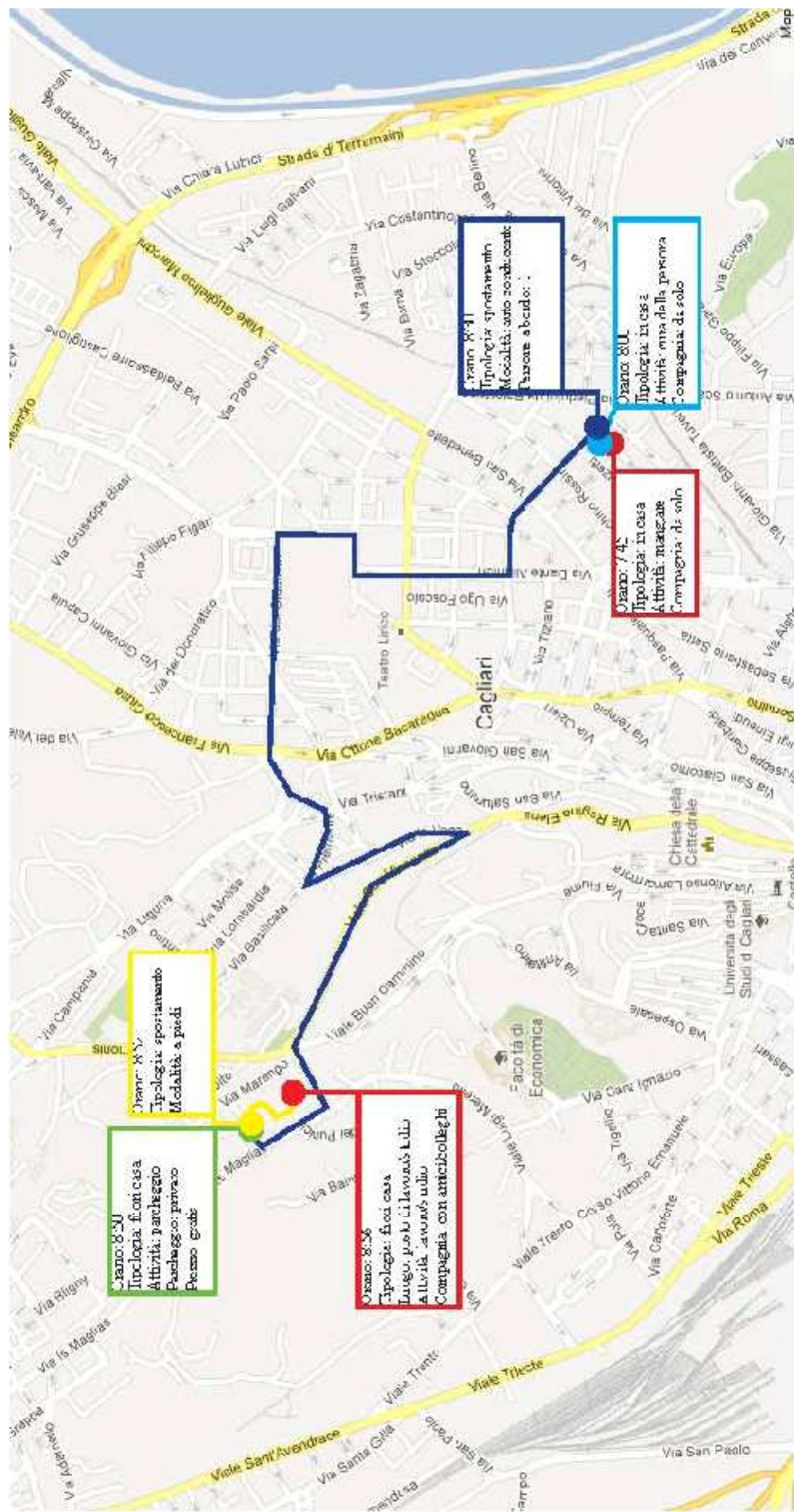


Figura 7 - Esempio output dati Activity Locator

## **4. TRATTAMENTO DEI DATI GPS**

La ricerca in esame, come detto, riprende il lavoro svolto da (Spissu, et al., 2011), che hanno effettuato un'indagine pilota utilizzando la medesima strumentazione (illustrata nel capitolo 3) su un campione di 26 utenti, i quali si differenziano dai dati attuali non solo per la quantità ma anche per la composizione, perchè per la prima indagine era stato selezionato un gruppo di studenti dell'Università di Cagliari che quindi non presentava le caratteristiche di eterogeneità proprie invece del database utilizzato in questa ricerca

L'obiettivo del lavoro presentato in questa sezione è quello di processare e analizzare i dati registrati attraverso l'Activity Locator. La prima fase dell'analisi ha riguardato la ricostruzione del database e l'applicazione di tecniche di map matching per associare ciascun punto GPS agli archi della rete. Ogni percorso mappato è stato poi verificato manualmente per eliminare eventuali incongruenze riscontrate durante la rielaborazione

La rete sulla quale sono stati elaborati i dati è quella estratta dal modello di macrosimulazione dell'area vasta di Cagliari, del quale si riporta, nei capitoli successivi, il riassunto del processo di calibrazione. Infatti, il modello sarà la base attraverso la quale saranno generati i dati necessari al compimento della seconda fase di questa applicazione, ossia il confronto tra i percorsi utilizzati per gli spostamenti avvenuti nell'ora di punta del mattino con quelli, appunto, simulati attraverso una piattaforma modellistica correttamente calibrata, per valutarne scostamenti e differenze. Attraverso un modello di scelta discreta, inoltre, sarà analizzata nel dettaglio l'intravariabilità degli spostamenti per ogni singolo utente, illustrata nel capitolo 4.5.

### **4.1. Il modello dell'area vasta di Cagliari**

Come anticipato precedentemente, le analisi sono state effettuate mediante l'ausilio di un software specialistico di simulazione del traffico. Infatti, il CRiMM (Centro di Ricerca Modelli e Mobilità dell'Università di Cagliari) ha implementato in CUBE (Citilabs), suite di programmi specialistici per la pianificazione dei trasporti, il modello di macrosimulazione dell'intera area vasta cagliaritana.

#### **4.1.1. Costruzione del grafo di rete**

La rete stradale dell'area vasta di Cagliari è stata implementata nel modello attraverso un grafo di rete. Quest'ultimo è stato aggiornato durante il corso degli anni ed è rappresentativo delle caratteristiche geometriche e funzionali della rete stradale di Cagliari e area vasta dell'anno 2012 ed è composto da:

- 606 zone origine destinazione (O/D) di cui 301 all'interno dell'area urbana di Cagliari, 198 comprese all'interno dell'area localizzata tra Cagliari e la SS 554 (comuni di Monserrato, Selargius, Quartu S. Elena, Quartucciu e la frazione di Pirri) e 107 esterne all'area delimitata dalla SS 554 e dal raccordo con la SS 195 ;
- circa 10.000 nodi;
- circa 22.500 archi.

Per ogni arco sono state definite le seguenti caratteristiche:

- nodo origine (a);
- nodo destinazione (b);
- distanza;
- velocità di percorrenza a flusso libero;
- tipologia;
- capacità teorica;
- curva di deflusso.

Sono state inoltre individuate e modellate 1273 intersezioni, di cui 111 semaforizzate, 22 intersezioni a livello sfalsato, 35 a rotatoria e 1105 intersezioni a raso regolate da segnale di stop o precedenza. La corretta modellazione degli archi e delle intersezioni è importante per la definizione della funzione di costo di ogni singolo arco, in quanto influisce sui ritardi accumulati alle intersezioni e di conseguenza sui tempi di percorrenza. La conoscenza di tale costo è una condizione essenziale per poter calcolare i percorsi possibili che collegano le diverse coppie origine/destinazione e di conseguenza poter modellare, in una successiva fase (assegnazione), i flussi di traffico sulla rete. Particolarmente importante nel calcolo del costo è la conoscenza del tempo di percorrenza dei vari archi nelle diverse condizioni di traffico. Esso è infatti variabile a seconda del grado di congestione della rete, andando da un valore minimo, dipendente dalla velocità di percorrenza in condizioni di flusso libero e dalla distanza, a valori sempre maggiori. Ad ogni arco viene quindi associata una curva di deflusso che descrive la variazione del tempo di percorrenza al variare del carico veicolare.

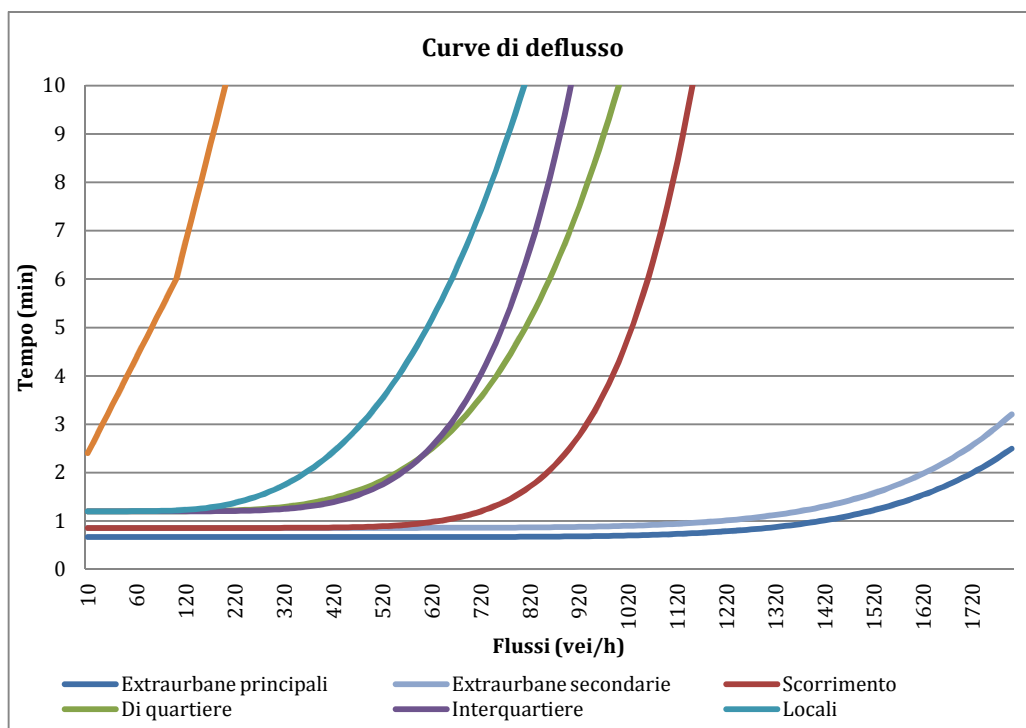


Figura 8 - Curve di deflusso (flussi -tempi)

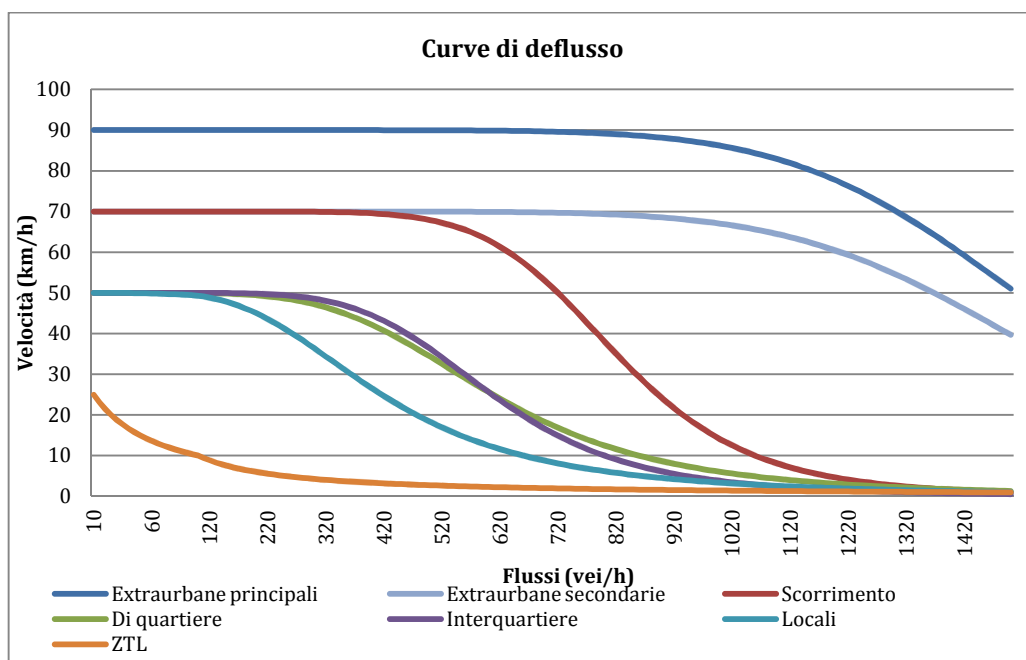


Figura 9 - Curve di deflusso (flussi -velocità)

Con l'obiettivo di definire adeguatamente la funzione preminente che ciascun elemento viario deve svolgere all'interno della rete stradale ed assegnare la corretta curva di deflusso, ad ogni arco è stata associato un attributo che ne identifica la gerarchia funzionale secondo la seguente classificazione:

- strade extraurbane principali;
- strade extraurbane secondarie;
- strade di scorrimento;
- strade interquartiere;
- strade di quartiere;
- strade locali;
- archi stradali ZTL;
- archi pedonali;
- corsie preferenziali autobus;
- ferrovia, metrotramvia.

La figura seguente illustra il grafo della rete stradale.

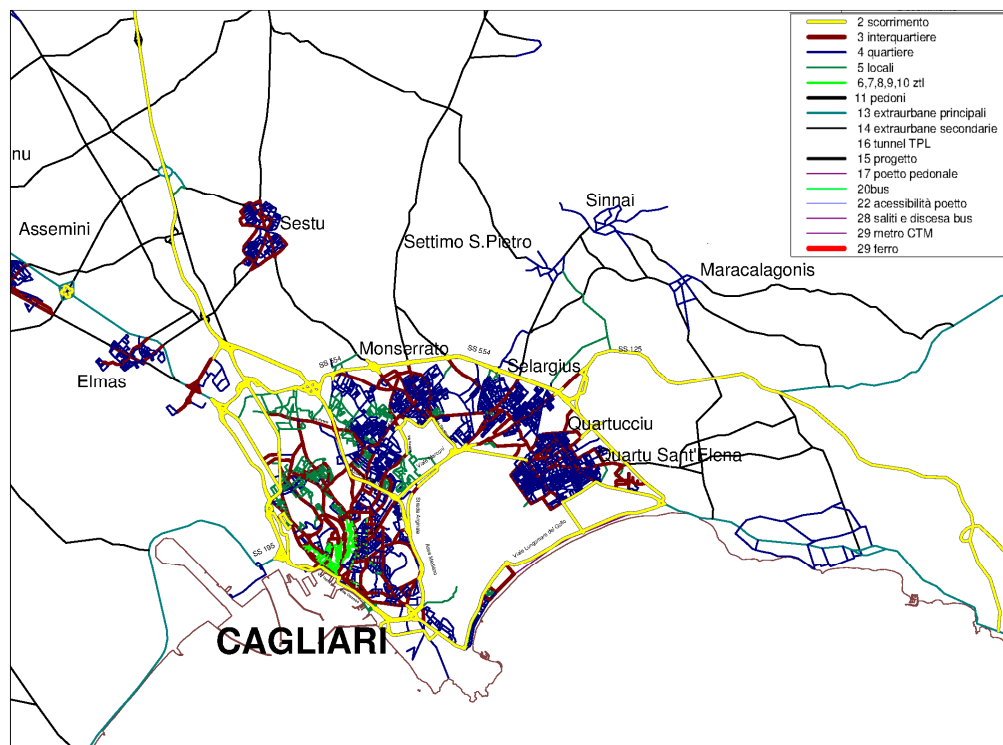


Figura 10 - Grafo di rete



#### **4.1.2. Il modello di domanda**

Il modello di domanda consente di mettere in relazione la quantità di spostamenti tra le diverse zone con le caratteristiche economiche, territoriali e sociali da cui dipende la domanda di viaggio. In particolare, per la stima della matrice è stato utilizzato un processo che aggiorna una matrice di partenza (Prior Matrix) sulla base dei rilievi di traffico.

La matrice originaria (1997/98), desunta da un'indagine origine e destinazione nell'area vasta cagliaritana, è stata aggiornata durante il corso degli anni utilizzando le diverse campagne di rilievi di traffico. Attualmente la matrice origine/destinazione è aggiornata al 2011, anno di realizzazione dell'ultima campagna di rilievi di traffico veicolare. Sono inclusi tutti gli spostamenti della fascia oraria di punta del mattino per tutti i motivi per i quali sono generati. In essa sono contenuti sia gli spostamenti pendolari (lavoro/studio) che quelli per altri motivi. L'aggiornamento al 2011 è avvenuto attraverso l'applicazione di un procedimento modellistico che consente di mettere in relazione e rendere coerenti gli spostamenti presenti nella matrice O/D con i conteggi di traffico eseguiti sulla rete, affinché gli stessi siano congruenti con i flussi di traffico sulla rete. Poiché i veicoli che transitano su ciascuna sezione interessano differenti coppie origine/destinazione, la ripartizione tra queste ultime avviene in proporzione al potenziale di attrazione e generazione di ciascuna zona oltreché ai valori presenti nella matrice di base (Prior Matrix). In definitiva quindi i dati in ingresso nel processo di stima della matrice sono:

1. conteggi veicolari (screen lines)
2. potenziali di attrazione/generazione delle zone (trip ends)
3. matrice di base (Prior Matrix).

Poiché i dati di base possono essere incongruenti tra di loro, in quanto provenienti da fonti e studi differenti, a ciascuno di essi viene associato un "livello di confidenza" che ne definisce la bontà e quindi il peso che ha nel concorrere alla determinazione dei valori della matrice. Tra le tre tipologie di dati sopra elencati, quelli che generalmente hanno il migliore livello di confidenza sono i conteggi di traffico, in quanto il valore rilevato è "certo". Quelli utilizzati per il processo di stima sono relativi ad un intervallo temporale che varia dal 2006 al 2013, periodo durante il quale sono stati rilevati i dati di traffico relativi a 186 sezioni stradali bidirezionali e 60 intersezioni (per un totale di circa 600 rilievi); il valore di confidenza è stato assegnato dando il valore più elevato ai dati più recenti (che risalgono a febbraio/marzo 2013), mentre per gli altri rilievi di traffico sono stati assegnati valori decrescenti in funzione del periodo temporale in cui sono stati effettuati. Per quanto riguarda i potenziali di generazione e attrazione, questi possono avere valori di confidenza molto alti se ricavati da conteggi posti al cordone della zona, o più bassi (rispetto ai conteggi di traffico) se ricavati da stime di analisi territoriale. Il livello di confidenza della matrice di

base ha invece un valore generalmente più basso rispetto a quello dei conteggi di traffico, poiché dipende dalla metodologia e dalla conoscenza di come essa sia stata stimata nel passato. Nel processo di stima, tuttavia, non è tanto importante il valore assoluto del livello di confidenza quanto quello relativo tra le componenti in gioco. Per questo motivo, generalmente l'individuazione dei livelli di confidenza avviene impostando inizialmente un valore, ritenuto elevato, a quelle informazioni che sono considerate più affidabili (solitamente i conteggi veicolari) e assegnando valori proporzionalmente più bassi a quelle informazioni che sono considerate meno affidabili.

La tabella seguente illustra i valori medi di confidenza adottati nel modello di Cagliari e area vasta.

<b>Descrizione</b>	<b>Valore di confidenza (%)</b>
Matrice base (Prior Matrix)	50
Conteggi di traffico (Screenlines)	80 – 100
Potenziali di generazione	70 – 90
Potenziali di attrazione	70 – 90

Tabella 6 - Stima della matrice ora di punta del mattino - valori di confidenza dei dati

<b>Domanda</b>	<b>Flussi veicolari</b>
7.30 – 8.30 (Stima 2010)	67.026
7.30 – 8.30 (Stima 2013)	71.514

Tabella 7 - Domanda di traffico veicoli leggeri (ora di punta del mattino 7.30-8.30)

<b>Domanda</b>	<b>Flussi veicolari</b>
7.30 – 8.30 (Stima 2010)	7.366
7.30 – 8.30 (Stima 2011)	7.409

Tabella 8 - Domanda di traffico mezzi pesanti (ora di punta del mattino 7.30-8.30)

Il dato complessivo di stima della matrice O/D tutti i motivi nella fascia 7.30 – 8.30 risulta pari a 71.514 vei/h per i veicoli leggeri e 7.409 vei/h per i mezzi pesanti, mentre quello della matrice o/d tutti i motivi costruita a partire dall'ultimo aggiornamento della matrice e aggiornata al 2013 risulta pari a 67.026 vei/h per i veicoli leggeri e 7.366 vei/h per i veicoli pesanti.

### 4.1.3. Il modello di assegnazione

Il modello di assegnazione consente di far interagire la domanda di trasporto (rappresentata dalla matrice O/D) con l'offerta di trasporto (grafo di rete) al fine di conoscere i flussi ed i costi sulla rete. In particolare, il modello di assegnazione ha la funzione di riprodurre i flussi di traffico attuali, effettuare valutazioni sulle criticità del sistema e costituire la base di confronto per la simulazione di eventuali alternative progettuali. Nei paragrafi successivi è sintetizzata la tecnica utilizzata per la costruzione del modello di assegnazione.

I flussi sono assegnati alla rete attraverso l'individuazione dei percorsi probabili che collegano le diverse coppie O/D sulla base di una funzione di costo che, nello specifico caso, è data dalla combinazione di due fattori: il tempo di percorrenza e il costo operativo del veicolo. L'enumerazione dei percorsi è basata sull'algoritmo di Dijkstra (1959), già illustrato nel capitolo 2.3.1. In sintesi:

$$C = \beta_1 T + \beta_2 C_{op} \quad (29)$$

dove T è il tempo, calcolato in funzione della velocità di percorrenza e della lunghezza d'arco, mentre  $C_{op}$  è il costo operativo, calcolato in base alla distanza imponendo 0,20€/km per i mezzi leggeri e 0,27€/km per quelli pesanti.

I coefficienti  $\beta_1$  e  $\beta_2$  rappresentano rispettivamente il peso del tempo e del costo operativo nella composizione del costo. I valori utilizzati valgono rispettivamente 1 per il coefficiente del tempo di viaggio e 0,4 per quello del costo operativo.

La tecnica utilizzata per il calcolo dei flussi assegnati è stata scelta quella del volume medio. Questa tecnica infatti garantisce la convergenza all'equilibrio di Wardrop (condizione teorica a cui si suppone tendano i flussi sulle reti di trasporto in assenza di fenomeni perturbativi) che è la condizione ottimale per la valutazione delle reti di trasporto.

L'assegnazione della matrice origine/destinazione attuale sulla rete esistente ha consentito l'estrapolazione di opportuni indicatori e la rappresentazione dei flussogrammi sul grafo di rete.

Gli indicatori calcolati sono:

- le distanze percorse dai veicoli (espresse in  $v_{ei} \cdot km$ )
- i tempi di percorrenza (espresi in ore)
- la velocità media (espressa in km/h)
- il grado di congestione (espresso tramite il rapporto  $v/c$ ).

Scenario		Velocità media	Distanza Percorsa	Tempi di Percorrenza
		km/h	Vei*km	Vei*h
Mattina	SDF	46,05	601.867	13.070

Tabella 9 - indicatori modello di assegnazione

Il rapporto v/c (volume di traffico/capacità stradale) è stato invece rappresentato graficamente nell'immagine riportata nelle pagine successive (nelle quali è indicato come q/c). Le figure Figura 12 e Figura 13 mostrano il flussogramma veicolare e il rapporto flusso/capacità.

#### 4.1.4. Calibrazione e Validazione del modello del traffico privato

Il processo di calibrazione del modello consiste, oltre che nella verifica puntuale delle caratteristiche degli archi e delle intersezioni, anche nell'ottenere che i risultati delle singole iterazioni del processo di assegnazione siano sempre costanti, stabili e convergenti (convergenza dell'assegnazione). La validazione del modello ha visto il confronto fra i flussi simulati e quelli rilevati sulla rete stradale durante i rilievi di traffico e il calcolo di alcuni indicatori statistici significativi. In particolare, è stato calcolato 1) l'errore medio assoluto MAPE, come somma delle differenze assolute degli scostamenti fra i valori osservati e quelli simulati rapportato al totale dei flussi osservati; 2) l'indice di correlazione tra i valori osservati e quelli simulati  $R^2$ ; 3) l'errore quadratico medio RMSE. I tre indicatori sono risultati pari rispettivamente al 18,7%, 0,92 e 22,9% per la simulazione dell'ora di punta del mattino, ampiamente all'interno delle soglie di calibrazione accettabili definite in letteratura.

Il grafico in Figura 11 illustra il confronto tra i flussi rilevati e quelli assegnati; ogni punto rappresenta una sezione di conteggio per il quale la posizione nel piano è individuata dal valore conteggiato (asse x) e dal valore assegnato (asse y). Nel caso di perfetta coincidenza tra il valore assegnato e il valore conteggiato il punto ricade sulla retta a 45°.

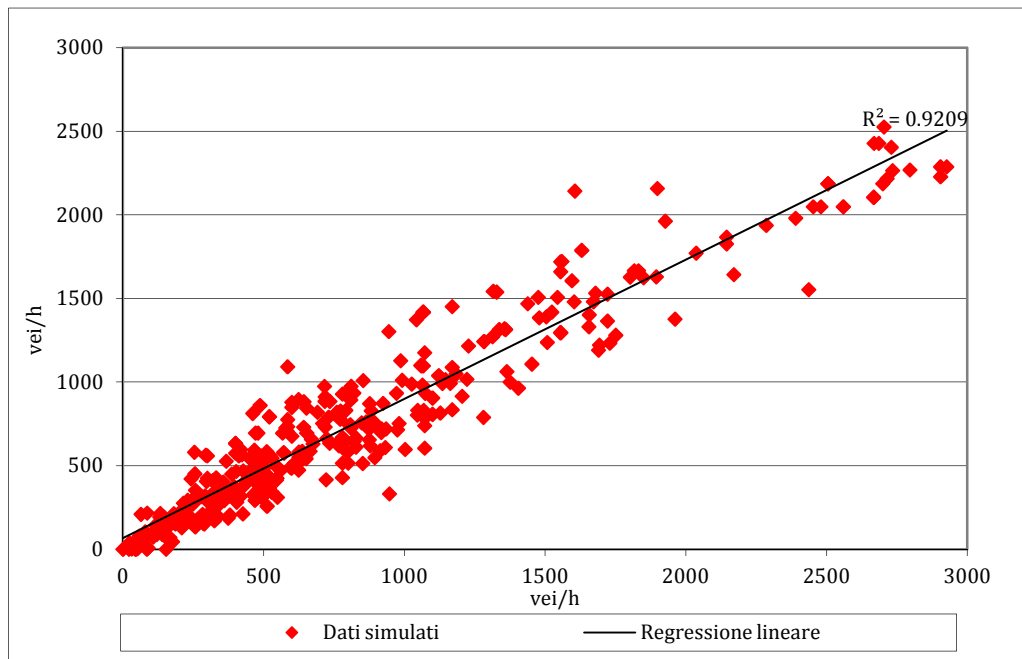


Figura 11 - Confronto flussi simulati e rilevati

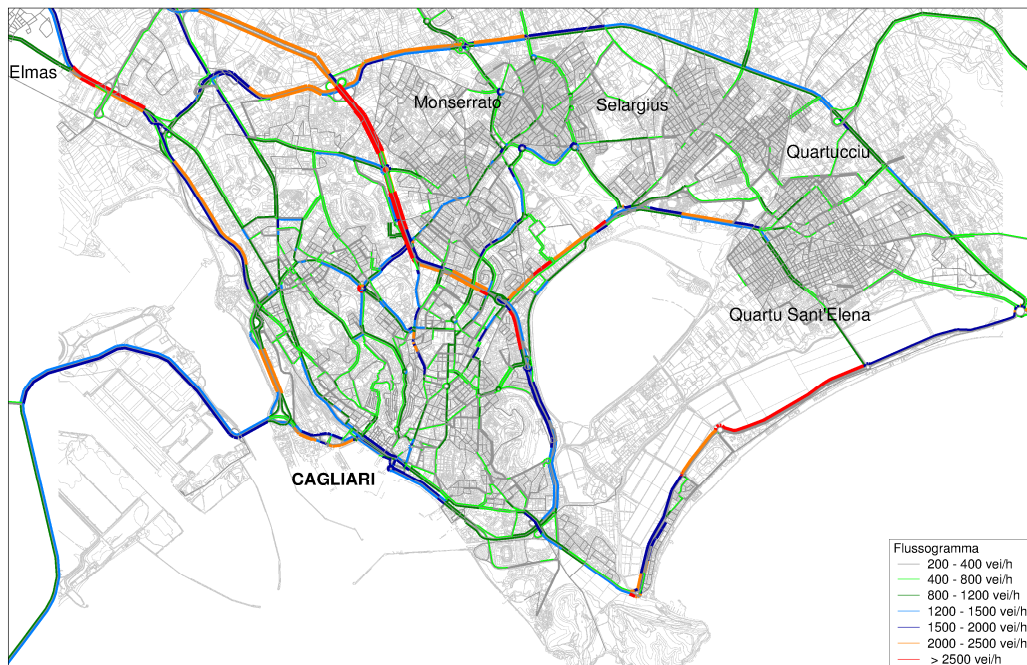


Figura 12 - Flussogramma ora di punta del mattino (7.30 - 8.30)

Il modello fornisce inoltre, per ogni arco stradale, il rapporto tra i flussi transitanti e la capacità (Figura 13). Questo indice viene utilizzato per individuare il livello di congestione e quindi le condizioni di flusso instabile, stabile e libero. Rapportando i dati di flusso alle capacità teoriche delle strade (rapporto  $q/c$  in figura), si può notare che gran parte delle infrastrutture viarie al cordone e interne ai centri abitati lavorano con valori compresi tra lo 0.5 e lo 0.75 (in colore verde), mentre in alcuni archi di registrano picchi prossimi alla capacità (in colore arancione e rosso 0.75 - 1). In particolare, gli archi viari più critici sono Viale Marconi, Viale La Plaia, alcuni tratti dell'Asse Mediano di Scorrimento, parte di Viale Poetto, Viale Colombo a Quartu, Via Italia a Pirri e Via Cesare Cabras a Monserrato. Per quanto riguarda il resto della rete, non si registrano condizioni particolarmente critiche che, con i tronchi stradali che lavorano in condizione di flusso prevalentemente libero o stabile, osservando infatti valori compresi tra 0.25 e 0.5.

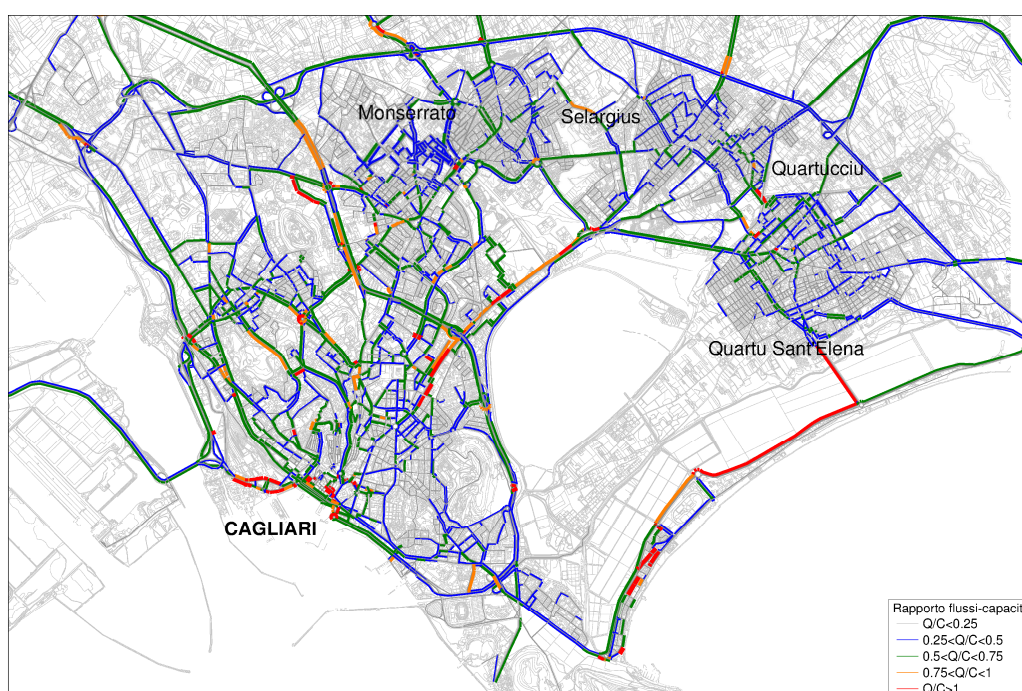


Figura 13 - Rapporto Flussi - Capacità, ora di punta del mattino (7.30 - 8.30)

## 4.2. Costruzione del database dei percorsi rilevati

La costruzione del database dei percorsi rilevati prevede la traduzione dei punti posizione GPS a percorsi veri e propri. Questa è una fase particolarmente importante, perché una scorretta associazione dei punti sugli archi (durante la fase di map matching) porterebbe ad avere un database di percorsi rilevati errato, comportando necessariamente degli errori nelle applicazioni in cui sarà utilizzato e limitando la validità delle analisi effettuate. Le informazioni di percorso (punti GPS) raccolte tramite l'Activity Locator sono registrate in un server sottoforma di uno shape file di punti (uno per ogni utente), nel quale ogni percorso è identificato da un identificativo (ID) e una sequenza ordinata di punti GPS, per ognuno dei quali si conoscono le informazioni riguardo gli istanti di arrivo e partenza, le coordinate spaziali ed altri attributi riferiti alle attività ad esso associate. Nel paragrafo seguente è illustrato brevemente il processo di map matching a cui sono stati sottoposti i dati.

### 4.2.1. Map matching

La procedura di map matching utilizzata in questo studio è quella a 2 stadi, basata sull'approccio geometrico che permette di tenere in considerazione gli attributi di viaggio registrati attraverso l'Activity Locator. Un algoritmo di minima distanza è implementato insieme ad un semplice settaggio che permette, in due fasi, di ridurre l'amplificazione dell'errore di allocazione GPS. In questo paragrafo si riporta brevemente la descrizione della metodologia applicata, mentre si rimanda alla lettura di Corona et al. (2012) per un maggiore approfondimento.

In prima istanza è stato necessario convertire il sistema di riferimento dell'intero insieme di punti posizione GPS, passando dal sistema geografico WGS84 a quello proiettato Gauss-Boaga, ossia quello relativo alla rete di Cagliari. Per considerare solamente i punti posti ad una distanza dagli archi consistente con l'errore GPS ammesso, è stato stabilito un raggio di 50 metri per ricerca dell'arco più vicino, non considerando quindi i nodi non compresi all'interno della misura considerata. Quelli restanti sono allocati all'arco più vicino secondo il metodo "point-to-curve matching" (White, et al., 2000).

Primo stadio: è stato utilizzato un algoritmo basato sulla distanza minima tra due punti successivi dello stesso spostamento, in modo da identificare il percorso dell'utente.

Secondo stadio: l'errore di tracciamento iniziale del GPS potrebbe essere amplificato dal fatto che alcuni punti potrebbero stazionare accanto ad archi non corretti. Di solito questo accade quando la rete urbana è molto densa. In questi casi è possibile la generazione di percorsi non logici. Siccome i punti posizione sono registrati con un intervallo abbastanza breve (5 secondi), due punti consecutivi potrebbero essere molto vicini e la distanza tra gli stessi

molto breve. Al fine di correggere i percorsi è stata implementata una semplice procedura che, nella simulazione, ignora i punti posizionati scorrettamente, sulla base della mutua distanza tra due punti consecutivi. Considerando la frequenza di campionamento, questa soglia è stata impostata a 50 metri, rimuovendo dal database i punti selezionati.

Tale metodologia ha permesso la ricostruzione del 95% dei percorsi utilizzati dagli utenti, che sono stati successivamente verificati singolarmente al fine della costruzione di un database consistente e solido.

#### **4.2.2. Elaborazione dei dati mappati**

Il risultato ottenuto è che ogni percorso “mappato” risulta costituito da un vettore di nodi consecutivi, identificati da un numero che corrisponde al nodo “fisico” all’interno del grafo di rete del modello. Mediante una successiva elaborazione è stato poi possibile convertire tale vettore in una sequenza del tipo A-B, che rappresentano rispettivamente il nodo iniziale e finale dell’arco considerato. Per le analisi con il modello di simulazione, per ogni percorso, sono state associate solo le caratteristiche di tempi e distanze, rimandando ad una fase successiva l’attribuzione delle caratteristiche qualitative. Per eliminare qualsiasi tipo di errore dovuto al map matching, inoltre, ogni percorso mappato è stato verificato manualmente ed eventualmente ricostruito in quanto, essendo il grafo di rete particolarmente denso, si riscontra spesso che durante il map matching un punto GPS sia associato ad un arco immediatamente prossimo a quello a cui effettivamente dovrebbe appartenere, ottenendo un risultato “visivamente” accettabile ma che non può essere utilizzato per l’implementazione nel modello di simulazione.

Infatti, questa tipologia di errore comporta che, al momento del calcolo del percorso, sia impossibile riprodurre quello osservato, perché i riferimenti con i quali è identificato sono errati. L’intero processo può essere schematizzato come di seguito:



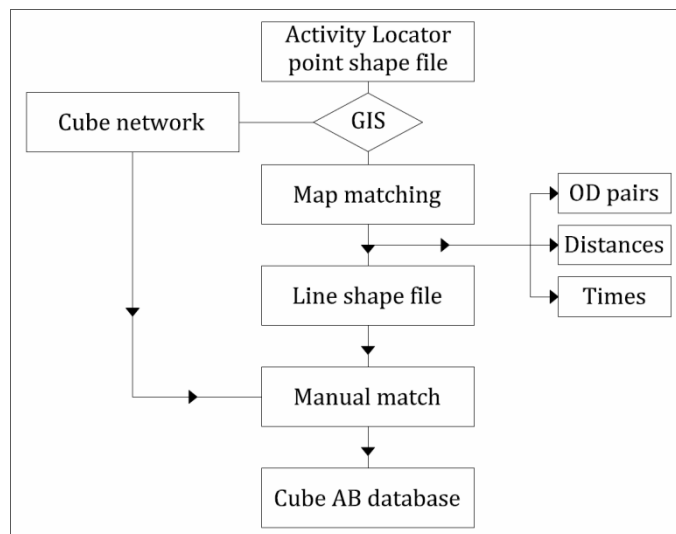


Figura 14 – Creazione del database dei percorsi rilevati

### 4.3. Costruzione del database dei percorsi simulati

Per la costruzione dei percorsi simulati è stata programmata, in CUBE, un'apposita applicazione il cui compito è quello di estrarre, per ogni osservazione, il corrispondente percorso di minimo costo. In particolare, sono estratti tre differenti percorsi minimi: i) minimo tempo; ii) minima distanza; iii) minimo costo (calcolato in funzione del tempo e della distanza). L'ultima tipologia è estratta per un numero di volte pari alle iterazioni necessarie al modello per raggiungere l'equilibrio.

Una volta predisposto il database relativo ai percorsi osservati, il passo successivo è stato quello di programmare in CUBE un'applicazione che permettesse, per ognuno di essi, di estrarre il corrispondente di minimo costo in termini di tempo o distanza e, inoltre, quello calcolato dal modello per ogni iterazione effettuata durante il processo di assegnazione (questo perché, se ci si fosse concentrati esclusivamente sul percorso all'equilibrio, si sarebbe commesso un errore, in quanto occorre tenere in considerazione tutti i percorsi che il modello calcola e utilizza per caricare la rete con i volumi di traffico).

Nel caso in esame, il modello ha raggiunto la convergenza in 70 iterazioni, quindi, per ogni coppia OD, è stato possibile estrapolare 72 percorsi simulati di cui: 1 di minima distanza, 1 di minor tempo, e 70 di minimo costo. Fra questi ultimi, molti sono calcolati ripetutamente, in quanto spesso accade che il modello consideri più volte lo stesso percorso come quello di minimo costo.

#### 4.4. Analisi dei dati nell'ora di punta del mattino

In questo paragrafo sono riportate le analisi del database dei percorsi avvenuti nell'ora di punta del mattino. L'obiettivo di questa fase dell'applicazione è quello di capire, rispetto a quanto studiato da altri autori in letteratura, come i dati in esame rispondano alle analisi da essi già effettuate. Una differenza rispetto a quanto esistente è rappresentata dal fatto che in questa applicazione è utilizzato un modello di macrosimulazione correttamente calibrato sulla rete reale, che dovrebbe fornire una base di confronto più robusta e realistica rispetto alle soluzioni GIS statiche delle ricerche precedenti. Allo scopo è stata programmata, in CUBE, un'applicazione apposita che permette di confrontare, dal punto di vista spaziale, i percorsi simulati con quelli osservati. Il funzionamento è il seguente: una volta inserito come input il percorso reale (ricostruito secondo le modalità descritte nei paragrafi precedenti) ed i nodi origine e destinazione, l'applicazione provvede ad individuare, dal database precedentemente costruito, tutti i percorsi che il modello ha calcolato tra quella specifica coppia OD. A questo punto viene generato un database di 70 righe, ciascuna di esse riferita ad un singolo percorso (tra la stessa coppia O/D) al quale sono associati gli attributi di costo, tempo e distanza. Attraverso il confronto spaziale tra i percorsi simulati e quelli osservati, si ottengono come risultato finale i km di lunghezza per cui il percorso osservato e quello simulato si sovrappongono perfettamente, requisito fondamentale per il calcolo della coverage.

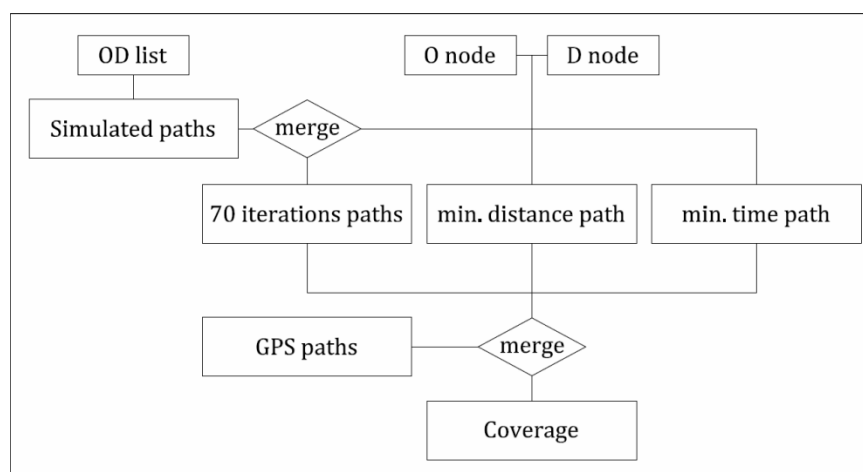


Figura 15 – Processo dei dati

Per l'applicazione presentata in questa ricerca sono stati scelti, dal campione totale di 8831 spostamenti, solamente quelli avvenuti in giorni feriali (dal lunedì al venerdì) nell'intervallo orario tra le 7:00 e le 10:00 del mattino, per la sola modalità auto-guidatore. Questa

restrizione è stata necessaria perché, come descritto nel 4.1, il modello è calibrato per una matrice riferita agli spostamenti dell'ora di punta del mattino. Per questo motivo è opportuno avere una base i cui dati provengano da condizioni di traffico corrispondenti a quelle simulate che, naturalmente, hanno particolare influenza nel calcolo del percorso di costo minimo. Non tutti i 109 utenti monitorati hanno però effettuato spostamenti con modalità auto-guidatore nell'intervallo orario selezionato, pertanto da una prima selezione ne sono stati individuati 93 aventi i requisiti necessari. Gli spostamenti totali riferiti a questo sottocampione sono 626, con una durata media di 18'42", una lunghezza media di 7.41 km, e una velocità media di circa 23.8 km/h. Per ogni spostamento è conosciuta la tipologia di attività che l'utente compie a destinazione, consentendo in tal modo di operare la distinzione classica tra spostamenti di tipo home-work e per attività discrezionali. La tabella di seguito mostra la ripartizione percentuale delle attività associate ai 626 spostamenti selezionati:

<b>Attività</b>	<b>%</b>
Lavoro/studio	73.56%
Acquisti/commissioni	9.13%
Salita/discesa passeggeri dall'auto	7.37%
Attività ricreative/sport	2.72%
Mangiare/preparazione pasti	2.24%
Pausa caffè	1.76%
Cura famiglia/casa	1.12%
Prepararsi/lavarsi	0.48%
Attesa mezzi/passaggio	0.48%
Volontariato	0.32%
Passeggiare	0.32%
Svago (tv, libri, internet)	0.16%
Visite mediche	0.16%
Dormire/relax	0.16%

Tabella 10 - Percentuali di attività

Come anticipato precedentemente, l'Activity Locator consente di conoscere anche la distanza degli spostamenti registrati, ottenuta attraverso l'interpolazione dei punti GPS. E' importante sottolineare il fatto che, nonostante i percorsi reali siano stati prima sottoposti a map-matching ed in seguito riverificati manualmente nel modello di simulazione, i due valori di distanza media differiscano solamente del 6.1%, ad evidenziare la bontà delle tecniche di analisi utilizzate. Infatti, facendo un confronto con la media delle distanze dei percorsi ottenuti direttamente dall'applicazione delle tecniche di map-matching, (senza quindi la post-ricostruzione descritta nel capitolo precedente), risulta che questa differisce da quella

GPS del 16.3%, ulteriore indicatore delle difficoltà che la metodologia ha incontrato nell'elaborazione dei dati.

#### *Coverage e deviazioni spazio-temporali*

Il primo confronto effettuato tra i percorsi reali e quelli simulati è stato quello che ha permesso di capire se gli utenti scelgono degli itinerari matematicamente considerati come di minimo costo, in funzione dei soli attributi di distanza e tempo di percorrenza. All'occorrenza, quindi, per ogni percorso è stata calcolata la coverage (percentuale di sovrapposizione) con il corrispondente di minimo costo determinato dal modello, secondo la seguente formula:

$$COVE_k^i = \frac{l_{sim}^i}{l_k} \quad (30)$$

dove  $i$  è l'indice relativo alla tipologia di percorso simulato (minimo costo, distanza, tempo) e  $k$  è l'identificativo dello spostamento,  $l_{sim}^i$  è la lunghezza per la quale i due percorsi a confronto condividono l'identico tracciato, mentre  $l_k$  è la lunghezza totale del percorso rilevato attraverso il GPS.

Data la natura dei dati a disposizione, uno stesso percorso può essere stato utilizzato dall'utente più volte durante il periodo d'indagine. È stata quindi effettuata un'ulteriore selezione, in modo tale da considerare, per i confronti, solo itinerari diversi tra loro (ad esempio, se un utente utilizza lo stesso percorso per 10 spostamenti, questo verrà considerato una volta sola). Il campione è stato quindi ridotto a 298 percorsi distinti (si è passati quindi da un database di spostamenti ad uno di percorsi). Il grafico successivo mostra come circa il 49% degli stessi abbia una coverage maggiore o uguale al 90%, mentre per la restante parte si registrano valori inferiori. Questo conferma ulteriormente quanto riporta la letteratura di settore, che dice che gli utenti, durante i propri spostamenti, non necessariamente utilizzano percorsi di minimo costo.

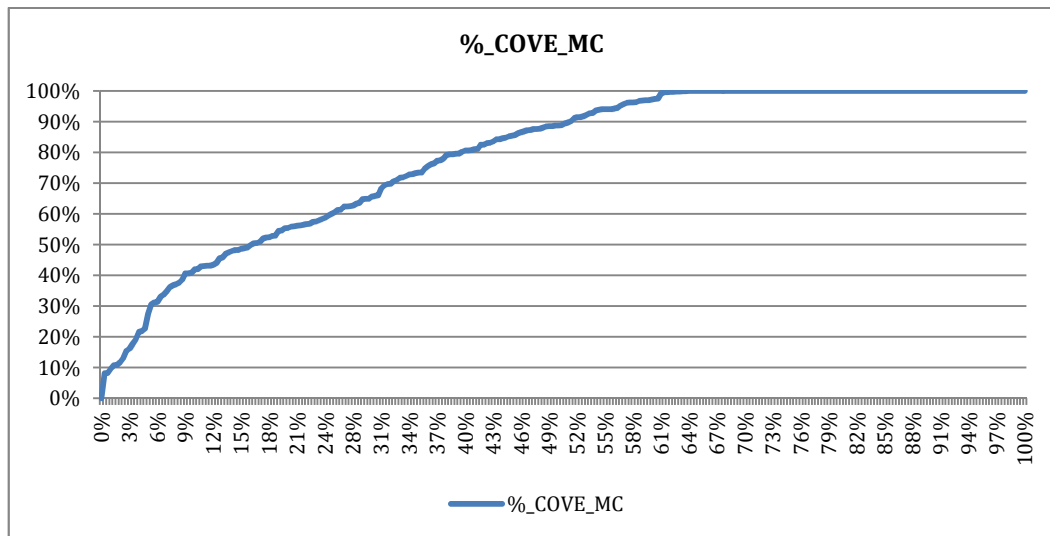


Figura 16 – Coverage del campione

Data inoltre la disponibilità dei tempi e delle distanze osservate, è stato possibile calcolare degli indici che rendono un'informazione circa la differenza, in termini percentuali, tra le distanze e tempi dei percorsi reali rispetto ai corrispondenti di costo minimo. Allo scopo, sono state utilizzate le formule suggerite da Spissu et al. (2011) relative al Time Deviation Index e al Distance Deviation Index, il cui scopo è quello di evidenziare quanto le scelte effettuate dagli utenti si discostino da quella "razionale" modellata.

*Distance Deviation Index* (DDI):

$$DDI = \frac{d_{obs} - d_{sim}}{d_{sim}} \quad (31)$$

dove  $d_{obs}$  è la lunghezza del percorso osservato e  $d_{sim}$  quella del simulato. Allo stesso modo, il *Time Deviation Index* (TDI) è:

$$TDI = \frac{t_{obs} - t_{sim}}{t_{sim}} \quad (32)$$

dove  $t_{obs}$  è il tempo di viaggio del percorso osservato e  $t_{sim}$  quello del simulato.

Il grafico successivo mostra gli indici calcolati in riferimento ai percorsi svolti per compiere attività di tipo lavoro/studio e per attività discrezionali.

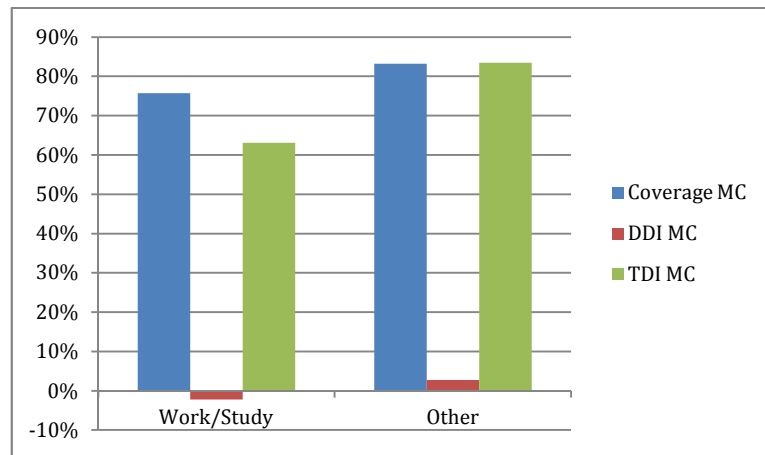


Figura 17 - Coverage, DDI and TDI divided by purpose

È possibile notare immediatamente come le deviazioni rispetto alle distanze siano particolarmente basse, ad indicare che gli utenti hanno una buona percezione spaziale dei propri spostamenti e della rete di trasporto, in quanto anche se le coverage medie indicano che i percorsi seguiti non corrispondono necessariamente a quelli di minimo costo (e quindi, matematicamente, più convenienti), la scelta di seguire un itinerario differente non porta ad eccessive differenze in termini di lunghezza. È invece interessante analizzare attentamente quello che si evidenzia guardando l'indicatore di deviazione del tempo (TDI). Infatti, mentre le coverage medie per le due tipologie di spostamenti sono quasi confrontabili (con scostamenti di pochi punti percentuale), le differenze in termini di TDI risultano maggiori. Questo fenomeno si può spiegare con un semplice ragionamento: il grafico mostra come gli utenti, in entrambi i casi, seguano dei percorsi di costo minimo per circa il 79% del totale; i valori minori di TDI per ciò che riguarda gli spostamenti casa/lavoro stanno ad indicare invece che, essendo questi ultimi tipicamente sistematici, spesso ripetuti nel tempo e caratterizzati da itinerari per i quali si presume che gli utenti abbiano acquisito notevole esperienza e consapevolezza, al momento della scelta del proprio percorso essi tendano a minimizzare non solo la distanza ma anche il tempo, al contrario di quanto avviene negli spostamenti per attività discrezionali.

La marcata differenza tra le deviazioni spaziali e temporali è attribuibile invece a diversi altri fattori. Mentre dal punto di vista della distanza è più verosimile avere a disposizione diversi itinerari simili (da cui i bassi valori del DDI), quando si considera il tempo di viaggio entrano in gioco altre dinamiche, tra le quali è possibile annoverare i fenomeni di congestione e i perditempo alle intersezioni semaforizzate. L'impatto di questi fattori di frizione può essere particolarmente importante specie per itinerari relativamente brevi, dove incrementi di

pochi minuti portano a delle variazioni percentuali elevate. tutto adeguata i tempi di viaggio. Quest'ultima spiegazione trova conferma nelle righe seguenti.

Il grafico successivo mostra la relazione tra coverage, TDI e DDI in funzione della lunghezza dello spostamento. Si ha la conferma che l'indice di deviazione rispetto alla distanza si mantiene più o meno costante su valori inferiori al 3%, mentre è interessante esaminare l'andamento del TDI. Infatti, assume valori particolarmente elevati per spostamenti di lunghezza inferiore ai 10 km, mentre per tutti quelli la cui distanza è maggiore le deviazioni non superano il 10%. Questo fenomeno può essere spiegato dal fatto che, per percorsi brevi, differenze di pochi minuti comportano deviazioni maggiori rispetto a quelle che si avrebbero per itinerari lunghi, dove le differenze tra i tempi sono compensate dal tempo di viaggio totale (ad esempio, 5 minuti di differenza su 10 minuti portano a un tDI=50%, mentre su 25 minuti sarebbe TDI=20%).

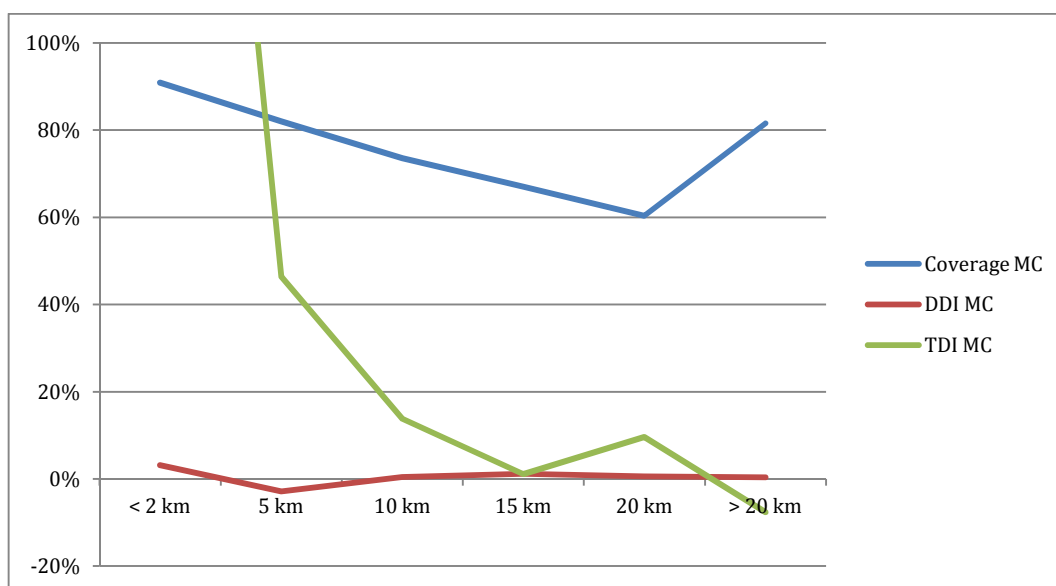


Figura 18 - Analisi in funzione della lunghezza dello spostamento

È interessante sottolineare il fatto che, se invece di considerare un solo spostamento tra una stessa coppia OD si effettuassero le analisi su tutto il campione, considerando quindi anche le ripetizioni dello stesso percorso da parte dei singoli utenti, gli indicatori presentano lo stesso andamento descritto in precedenza, fatta eccezione per il TDI degli spostamenti lavoro/studio, che mostra un valore del 20%. Questo è un'ulteriore conferma di quanto sottolineato, ossia per spostamenti sistematici e abitudinari l'esperienza acquisita porti gli utenti ad effettuare delle scelte che tengono conto maggiormente degli attributi tempo e distanza rispetto ad altri fattori.

In conclusione, I risultati mostrano come, in generale, il modello riproduca circa il 78% dei percorsi attualmente scelti dal guidatore, in linea con quanto riportato da in letteratura da (Ortuzar & Willumsen, 2001), che indicano che le funzioni di costo basate sul tempo di viaggio e distanza riescono a replicare circa il 60%-80% dei percorsi osservati nella realtà. È stato trovato che la distanza sia l'attributo che tenda maggiormente ad essere minimizzato, indice di una buona percezione spaziale, mentre lo stesso non accade relativamente ai tempi di viaggio. Per spostamenti abituali di tipo casa-lavoro, dove il guidatore ha familiarità con il percorso e le condizioni del traffico, la scelta del percorso è molto simile a quella definibile come razionale, dove inoltre gli indici di deviazione mostrano valori più bassi rispetto a quelli per spostamenti relativi ad attività discrezionali.

#### 4.5. Intravariabilità

Un'ulteriore attenzione è stata dedicata agli spostamenti (di uno stesso utente) ripetuti tra la stessa OD, consentendo quindi di valutare il grado di intravariabilità dei percorsi seguiti. Sul totale di 626 spostamenti già analizzati nel precedente paragrafo, 381 presentano le caratteristiche richieste, pari a circa il 60.9%. Il grafico seguente mostra la percentuale del campione che ripete il percorso tra la stessa OD di un numero di volte pari a quanto indicato nell'asse delle ascisse.

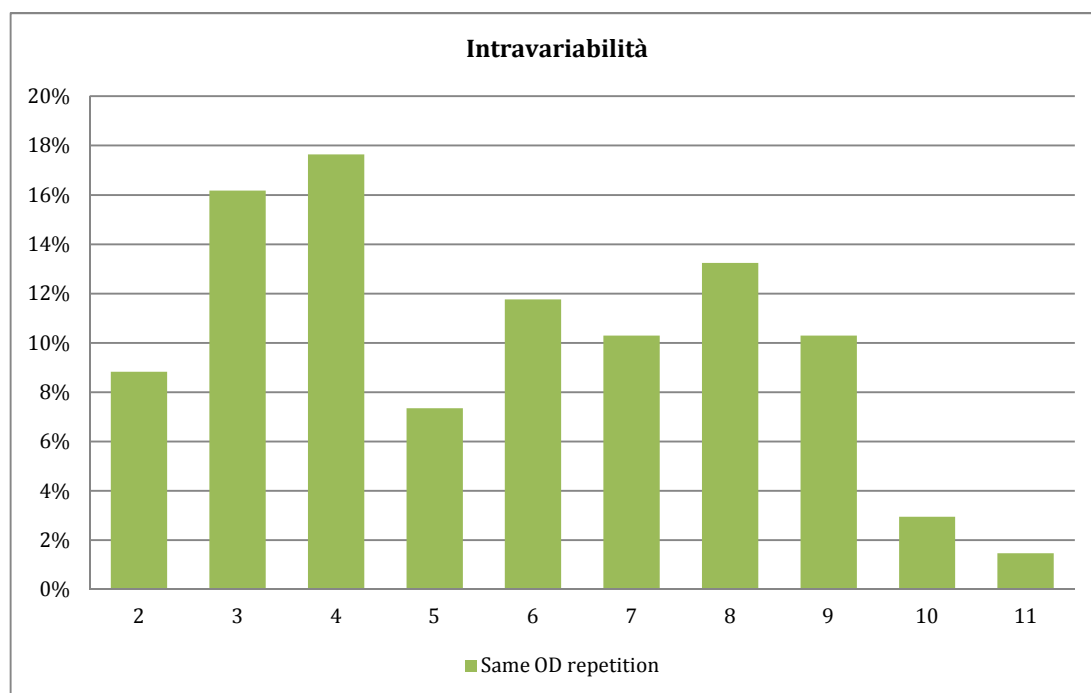


Figura 19 – Ripetizione dello stesso spostamento per utente-coppia O/D



Risulta che il 69,1% degli utenti usi sempre lo stesso itinerario, il 22,1% almeno due mentre il restante 8,8% sceglie tre o più percorsi. Nell'arco delle due settimane di indagine, gli utilizzatori di un solo percorso compiono in media 5.28 spostamenti tra la stessa OD, contro i 6.29 di quelli che scelgono più itinerari. Concentrando l'attenzione su questa frazione del campione, risulta un percorso prevalente che l'utente sceglie più spesso rispetto agli altri, con una percentuale media del 68%.

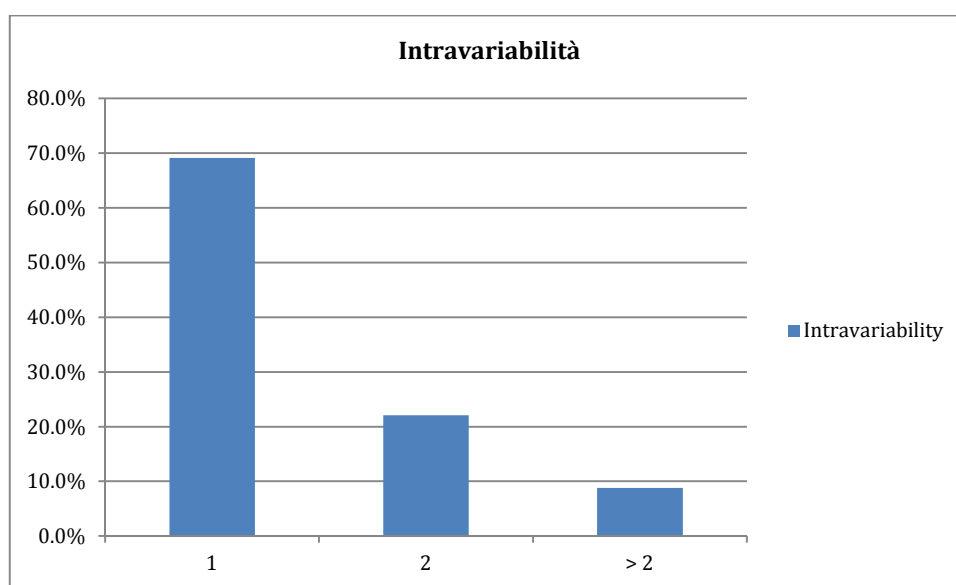


Figura 20 - Intravariabilità

È utile quindi verificare come ed in che misura il percorso più utilizzato si differenzi rispetto agli altri, utilizzando come base di confronto gli indici calcolati per le analisi effettuate rispetto ai percorsi di minimo costo. I risultati mostrano che, nei casi in cui il più usato (49 spostamenti) ha un coverage maggiore, nel 98% delle volte l'istante di partenza risulta successivo a quello del secondo migliore, in media di circa 13 minuti (8:29 contro 8:16). Partendo in ritardo, la rete risulta meno congestionata, in quanto dalle 8:00 alle 8:30 si registrano particolari difficoltà dovute agli ingressi degli studenti a scuola e alla maggiore frequenza dei mezzi di trasporto pubblico che, viaggiando perlopiù in sede promiscua, costituiscono un'impedenza particolarmente importante al libero deflusso dei veicoli. Nei 20 casi in cui il percorso scelto più di frequente si discosta maggiormente rispetto a quello dalle caratteristiche di minimo costo, gli orari di partenza non presentano un andamento regolare. Praticamente non c'è una ripetizione di una stessa caratteristica che, come nel caso precedente, poteva essere rappresentata dal fatto che gli spostamenti aventi una coverage maggiore iniziassero sempre in un momento successivo.

É interessante confrontare i risultati ottenuti con quelli dello studio pilota effettuato in precedenza da (Spissu, et al., 2011), dove per spostamenti ripetuti tra la stessa OD il 92,7% degli utenti usa solo un percorso, mentre i restanti due o più. La differenza, circa il 23%, con il presente studio può essere attribuita al fatto che si abbiano considerato, in questo caso, solo i percorsi effettuati durante l'ora di punta del mattino per un campione eterogeneo di utenti, mentre lo studio pilota era relativo all'intera giornata. Tuttavia, altri studi relativi a questo campo di ricerca hanno riportato diverse percentuali: ad esempio, Jan, et al. (2000) trovano che il 34% degli utenti usa lo stesso percorso; Li, et al. (2005) trovano un valore di 40%, mentre il 60% almeno due. Zhu & Levinson (2010) invece trovano che il 40% usino più di un percorso, con una differenza di circa il 10% rispetto a quelle evidenziate in questo lavoro (31%), attribuibile per lo più alle diverse caratteristiche del campione utilizzato.

Per valutare le dinamiche legate all'intravariabilità, si è deciso di sviluppare un modello di scelta discreta, basandosi sulla comune teoria dell'utilità casuale (Domencich & McFadden, 1975; Williams, 1977). La specificazione utilizzata è quella del mixed logit, ML, (Ben Akiva & Bolduc, 1996; McFadden & Train, 2000), per tenere conto di due aspetti: l'effetto panel, dovuto al fatto che i dati si riferiscono ad utenti che hanno compiuto più scelte nell'arco del periodo di rilievo, ragione per cui si avranno a disposizione informazioni che differiscono tra loro esclusivamente per il tempo di percorrenza e gli istanti di partenza e arrivo; la seconda è la differenza di percezione, tra gli utenti, del peso da attribuire ad alcune variabili particolarmente sensibili e di interpretazione soggettiva (ad esempio il tempo). Queste tipologie di modelli sono risultate particolarmente efficaci nello studio dell'intravariabilità, come riportato ad esempio nei lavori di Han, et al. (2001), Srinivasan & Mahmassani (2003) e Ben-Elia & Shiftan (2010), di cui si è discusso precedentemente.

La variabile dipendente rappresenta la caratteristica di abitudine o meno del percorso, dove per abituale si intende quello che, durante il periodo di rilievo, è stato utilizzato dallo stesso utente il maggior numero di volte negli spostamenti tra la stessa coppia OD. Non si tratta infatti di un modello di scelta del percorso convenzionale (che include l'insieme di scelta delle alternative), ma bensì l'obiettivo finale è quello di analizzare, per mezzo di un modello logit binario (regressione logistica), quali siano le caratteristiche socio-economiche dell'utente e quelle oggettive del percorso e che fanno sì che questo sia considerato abituale.

#### **4.5.1. Il modello mixed logit**

La struttura di base del ML è uguale a quella di tutti i modelli di scelta discreta tipicamente utilizzati nel settore dei trasporti, con l'idea innovativa di considerare il termine aleatorio formato da più di una componente, di cui almeno una  $\varepsilon_{jq}$  distribuito IID Gumbel ed un'altra

$\eta_{jq}$  che può assumere qualunque distribuzione e che consente di modellare numerosi fenomeni quali correlazione tra alternative, correlazione tra utenti, eteroschedasticità, eterogeneità casuale nelle preferenze, etc. Il modello ML ha pertanto la seguente specificazione:

$$U_{jq} = \theta x_{jq} + \eta_{jq} + \varepsilon_{jq} \quad (33)$$

Poiché si ha una componente additiva che si distribuisce IID Gumbel, integrando rispetto a  $\varepsilon_{jq}$  si ottiene la forma classica di un Logit Multinomiale:

$$L_{iq}(\eta) = \frac{\exp(\theta x_{iq} + \eta_{iq})}{\sum_j \exp(\theta x_{jq} + \eta_{jq})} \quad (34)$$

Tuttavia, poiché  $\eta_{jq}$  non è nota, la probabilità di scelta è una probabilità condizionata. Assumendo per  $\eta_{jq}$  una qualunque distribuzione, con funzione di densità  $f(\eta|\Omega)$  dove  $\Omega$  sono i parametri fissi della distribuzione (media e deviazione standard), la probabilità di scelta non condizionata è data dall'integrale di  $L_{jq}$  su tutti i valori  $\eta_{jq}$  pesati dalla rispettiva funzione di densità  $f(\eta|\Omega)$ :

$$P_{jq} = \int L_{jq}(\eta) f(\eta|\Omega) d\eta \quad (35)$$

La maggiore flessibilità del ML, che non presenta il problema della IIA (indipendenza delle alternative irrilevanti) e che consente una simulazione più reale dei comportamenti degli utenti, si riflette tuttavia in una maggiore complessità nella stima dei parametri. Infatti, a differenza di quanto si verifica per il MNL, la probabilità ML non presenta una forma chiusa e la stima dei parametri richiede la risoluzione dell'integrale multiplo attraverso metodi basati sulla simulazione. Le due forme più comuni del ML sono quella a coefficienti casuali (random coefficient, RC) e quella con componente d'errore (error component, EC), che differiscono tra loro a seconda del tipo di attributo a cui è associato  $\eta_{jq}$ . Nel modello ML-RC gli attributi associati a  $\eta_{jq}$  sono noti al modellista e la specificazione è la seguente:

$$U_{jq} = \theta_q x_{jq} + \varepsilon_{jq} \quad (36)$$

in cui l'errore è IID Gumbel e dove si riconosce che il parametro di preferenza non è fisso ma varia in funzione della popolazione. Senza perdita di generalità, si può scrivere che:

$$\theta_q = \theta + \eta_q \quad (37)$$

Quindi

$$U_{jq} = \theta x_{jq} + \eta_q x_{jq} + \varepsilon_{jq} \quad (38)$$

dove la correlazione tra le alternative è data dalla presenza del termine comune  $\eta_q$ . Il modello ML-EC, invece, ha la seguente specificazione:

$$U_{jq} = \theta_q x_{jq} + \eta_{jq} + \varepsilon_{jq} \quad (39)$$

che fondamentale è un modello con coefficienti fissi in cui il termine d'errore ha due componenti, dove si assume che il termine  $\eta_q$  si distribuisca con media 0. Ulteriori informazioni sui modelli logit misti è possibile ricavarle in Train (2009).

A partire dalle basi teoriche illustrate sono stati sviluppati degli appositi codici in grado di specificare, nel linguaggio del software GAUSS, i diversi modelli.

#### 4.5.2. Caratteristiche dei dati

Per le analisi modellistiche, solo 361 spostamenti, riferiti a 66 utenti, hanno le caratteristiche di ripetitività (per motivi casa-lavoro), dove il 29% degli utenti usa più di un percorso tra la stessa coppia O/D. La lunghezza media degli spostamenti è di 8.13 km con un tempo di viaggio pari 18.48 minuti, equivalenti a circa 3 minuti a km. Lo 0.3% del viaggio avviene su strade congestionate ( $V/C > 0.8$ ), mentre lo 0.3% e 0.4% avvengono su strade a scorrimento veloce e interquartiere, rispettivamente. Il campione è composto dal 45.5% di maschi e 54.5% di femmine, con età compresa tra 31 e 40 anni (39.4%), 18 e 30 (28.8%), 41 e 60 (31.8%). Il 60.6% sono impiegati, il 25.8% liberi professionisti, 10.6% studenti e il restante disoccupati. Il livello di educazione predominante è la laurea (39.4%), seguita dal diploma (30.3%). Il 43.9% degli utenti ha un reddito mensile medio compreso tra i 1000€ e i 2000€, con il 24.2% che guadagna meno di 1000€ al mese. Circa il 50% degli utenti guida per meno di 15000 km all'anno, mentre il 22.7% e il 15.2% percorrono, rispettivamente, tra i 15000km e i 25000 e più di 25000km all'anno. Una gran parte (43.9%) dichiara di spendere tra i 50€ e i 100€ al mese per i trasporti, mentre il 37.9% tra 100€ e 300€. Il 30.3% appartiene ad un gruppo familiare composto da 4 membri, il 24,2% da 2 e il 22.7% da 3 membri, con il 31.8% che ha almeno un figlio. Il 51.5% ha due auto, il 21.2% tre e il 19.7% solo una, tra le quali la più posseduta è un'utilitaria (74.2%). Questi dati sono riassunti nella tabella seguente, che mostra tutte le caratteristiche di questo sottocampione.

<b>ROUTE</b>	<b>N°</b>	<b>Ave.</b>	<b>Std. Dev</b>	<b>USER</b>	<b>N°</b>	<b>%</b>
Average V/C ratio		0.62	0.21	<b>Age group</b>		
Percentage of route where V/C > 0.8		0.30	0.21	18 - 30	19	28.8%
Number of traffic lights per km		0.56	0.58	31 - 40	26	39.4%
Percentage of route consisting of highways		0.33	0.28	41 - 60	21	31.8%
Percentage of route consisting of inter-district roads		0.36	0.27	61 - 80	0	0.0%
Travel time in minutes		18.48	10.77	<b>Gender</b>		
Route length		8.13	7.01	Male	30	45.5%
Least cost set	153			Female	36	54.5%
<b>TRIP</b>				<b>Occupation</b>		
Difference in departure time vs average		0.02	12.69	Student	7	10.6%
Number of trips started in advance	46			Employee	40	60.6%
Number of trips started late	42			Self-employed	17	25.8%
				Unemployed	2	3.0%
<b>HOUSEHOLD</b>	<b>N°</b>	<b>%</b>		<b>Education</b>		
<b>Children</b>	21	31.8%		High school graduation	20	30.3%
<b>Household members</b>				Professional specialization	4	6.1%
1	9	13.6%		Degree	26	39.4%
2	16	24.2%		Post Degree	14	21.2%
3	15	22.7%		Primary school	2	3.0%
4	20	30.3%		<b>Monthly income</b>		
5	6	9.1%		Less than €1000	16	24.2%
<b>Number of cars per household</b>				€1000 - €2000	29	43.9%
1	13	19.7%		€2000 - €4000	12	18.2%
2	34	51.5%		More than €4000	4	6.1%
3	14	21.2%		Don't know	1	1.5%
4	3	4.5%		Don't have	4	6.1%
5	2	3.0%		<b>Km/Year</b>		
<b>Car type</b>				Less than 15,000 km	33	50.0%
Sedan	8	12.1%		15,000 - 25,000 km	15	22.7%
Two seater	5	7.6%		More than 25,000 km	10	15.2%
Station Wagon - minivan	4	6.1%		Don't know	8	12.1%
Economical	49	74.2%		<b>€/month for transport</b>		
				Less than €50	8	12.1%
				€50 - €100	29	43.9%
				€100 - €300	25	37.9%
				More than €300	2	3.0%
				Don't know	2	3.0%

Tabella 11 - Caratteristiche del campione

### **4.5.3. Caratteristiche delle variabili**

Sono state considerate due differenti tipologie di variabili: del livello di servizio e socio-economiche. Le prime sono state oggetto di una ulteriore suddivisione in variabili del percorso e dello spostamento, mentre le seconde in due gruppi riferiti all'individuo e alla famiglia. Considerando il livello di servizio, la differenza (in minuti) tra il tempo di partenza del singolo spostamento e quello medio di tutti gli spostamenti dell'utente (considerando ovviamente solo quelli del sottocampione selezionato), la media del rapporto  $V/C$ , la percentuale di percorso congestionata ( $V/C > 0.8$ ), il numero di semafori per km, la percentuale del percorso costituita da strade a scorrimento veloce o interquartiere, il tempo di viaggio, la distanza e i minuti per km di spostamento sono tutte considerate come continue, mentre la particolarità di partire in anticipo o in ritardo rispetto alla media (vedi sopra) e l'appartenenza del percorso a quelli di minimo costo sono considerate come dummies. Le variabili socioeconomiche sono state selezionate utilizzando come base le ricerche precedenti, alcune assunzioni personali e da informazioni fornite dagli utenti durante la compilazione dei questionari, strutturate in maniera tale da tenere conto di qualsiasi differenza nelle preferenze dei diversi gruppi di persone. Tutte le variabili legate all'utente sono considerate come dummies, in accordo con le categorie mostrate nella tabella precedente, così come i figli e la categoria veicolare, mentre le altre sono considerate come continue.

In aggiunta alle variabili elencate, in letteratura sono presenti altre variabili considerate particolarmente importanti, come ad esempio il numero di stop intermedi e l'affidabilità del tempo di viaggio. Ad esempio, Li et. al (2005) hanno mostrato che la propensione a scegliere più percorsi tra la stessa coppia O/D aumenta proporzionalmente al numero di fermate intermedie lungo il tragitto, mentre l'affidabilità del tempo di viaggio è usata spesso nelle ricerche riguardanti la scelta del percorso. Sfortunatamente, il sottocampione a disposizione non permette il calcolo di queste variabili. È intenzione dell'autore, in eventuali ricerche future, testare queste variabili in casi di utilizzo dell'intero database di dati, dove saranno disponibili tutte le informazioni di percorso.

### **4.5.4. Risultati**

Sono state specificate differenti specificazioni modellistiche per identificare quali siano le variabili più significative. I risultati mostrano come il numero di semafori per km e la percentuale di strade a scorrimento veloce abbiano particolare rilevanza statistica, con segno rispettivamente negativo e positivo. Questo conferma l'ipotesi che gli utenti tendano a minimizzare il numero di semafori incontrati durante il tragitto e preferiscano strade più

scorrevoli rispetto a quelle locali, coerentemente con quanto trovato anche da Zhu and Levinson (2010). Anche l'appartenenza all'insieme dei percorsi matematicamente di costo minimo presenta notevole significatività, un risultato che è in linea con quanto mostrato nella precedente fase della ricerca (analisi dei percorsi attraverso il modello di simulazione dell'area vasta di Cagliari). Sono interessanti ma non significativi i segni delle variabili associate all'istante di partenza in anticipo/ritardo rispetto alla media, rispettivamente positivo e negativo, ad indicare che quando gli utenti escono da casa in ritardo tendano ad utilizzare il proprio percorso abituale, probabilmente perché preferiscono affidarsi ad un percorso conosciuto e con un tempo di viaggio più o meno certo per arrivare in tempo a lavoro. Lo stesso vale per il segno (positivo) associato alla differenza, in minuti, dell'istante di partenza, che conferma la preferenza per il percorso "solito" quando si esce da casa in ritardo. Il tempo di viaggio e la distanza non risultano particolarmente significativo, così come minuti di viaggio per km, probabilmente a causa delle esigue dimensioni del campione utilizzato. L'ultima ha inoltre un valore positivo, suggerendo che gli utenti tendano ad utilizzare percorsi più lenti, in contrasto con quanto ci si aspetterebbe in prima ipotesi, ma confermato dal segno positivo della variabili associate al grado di congestione ( $V/C$  medio e percentuale di percorso con  $V/C > 0.8$ , entrambe non significative). Una spiegazione è possibile trovarla nella scarsa conoscenza delle alternative a disposizione, oppure nell'impossibilità di effettuare delle scelte differenti (una sola alternativa disponibile). Anche Srinivasan and Mahmassani (2003) hanno notato lo stesso comportamento, giustificandolo nella difficoltà di re-routing nella rete. Ancora, gli effetti dell'abitudine e familiarità possono giocare un ruolo importante nella scelta del percorso, portando gli utenti a preferire percorsi lenti e congestionati ma "sicuri", piuttosto che altri teoricamente più convenienti. Due altre variabili che non sono risultate significative sono la percentuale di strade interquartiere e l'appartenenza all'insieme di percorsi di minima lunghezza, entrambe con segno negativo. La spiegazione della seconda è identificabile nel fatto che l'utente percepisca molto bene le caratteristiche del percorso e come risultato tenda a scegliere quello più corto. Le strade interquartiere, invece, descrivono una grossa percentuale della rete urbana di Cagliari, quindi è possibile che anche scegliendo un percorso alternativo si percorra lunghi tratti con queste caratteristiche. È stata inoltre stimata la variabilità di percezione della stessa variabile tra differenti utenti, trovando che solo i minuti per km siano significativi (utilizzando una distribuzione normale standard), chiaro segno di come gli utenti interpretino diversamente questo valore e ulteriore dimostrazione dell'importante ruolo che la percezione individuale occupa nei processi decisionali. Tra le variabili legate all'età, il gruppo più significativo è quello tra i 18 e 30 anni, con segno positivo, indicatore che questo gruppo di utenti tenda ad utilizzare il percorso abituale. È interessante notare come il segno del gruppo 31-40 sia negativo, indicando che questo gruppo abbia una maggiore

propensione ad usare più percorsi, probabilmente perché costituito da guidatori con più esperienza. Tra le variabili relativi al genere, si riporta che i maschi tendano a utilizzare più percorsi, con inoltre buona significatività statistica, probabilmente per la tendenza (o stereotipo) che abbiano più esperienza nel guidare, in accordo con quanto trovato da (Khattak, et al., 1995; Mannering & Kim, 1994; Han, et al., 2001; Xu, et al., 2010). tra le altre categorie di variabili socio-economiche, non é stata trovata particolare rilevanza statistica per l'occupazione, il numero di km percorsi all'anno, reddito, spesa mensile per i trasporti e classe della vettura. Sono state inoltre testate alcune variabili d'interazione, per valutare l'influenza di particolari attributi del livello di servizio del sistema su particolari categorie di utenti. É stato trovato che, in relazione con l'aumento del tempo di percorrenza per km, gli utenti che percorrono tra i 15000 e i 25000 km all'anno e quelli con reddito superiore a 4000 € mensili abbiano maggiore propensione a cambiare percorso. Per i primi la spiegazione si trova nella maggiore conoscenza della rete e delle possibili alternative a disposizione, vista l'esperienza acquisita "sul campo", mentre per i secondi il segno é associabile al maggiore valore che questa categoria associa al tempo. La tabella seguente riassume brevemente i risultati elencati, riportando le stime e i valori di un modello nel quale sono state implementate solamente le variabili per le quali é stata riscontrata una buona significatività statistica.

<b>Variable</b>	<b>Model</b>	
	<i>Coeff</i>	<i>t-test</i>
<b>Level of service</b>		
Number of traffic lights per km	-1.4516	-3.644
Percentage of route consisting of highways	4.293	3.985
Sigma minutes per km	0.3684	2.093
Least cost set	2.712	3.85
<b>Socio-economic</b>		
Age from 18 to 31	1.6516	2.765
Males	-1.1623	-1.442
<b>Interaction</b>		
Minutes per km for 15000-25000 km	-0.3493	-1.671
Minutes per km for more than €4000 /month	-0.6242	-1.481
<b>Constant</b>		
Observations	1.7667	2.966
Users		361
Final Log Likelihood		66
		-15.987972

Tabella 12 - Risultati del modello



## 5. STIMA DI UN MODELLO DI SCELTA DEL PERCORSO

Come argomentato nel capitolo 2.2, il comportamento di scelta del percorso è generalmente inquadrato come un processo a due stadi: generazione dei percorsi alternativi e modellazione della scelta dall'insieme generato. Nonostante la generazione dell'insieme di scelta possa essere sia implicita che esplicita, per diversi vantaggi teorici e computazionali la seconda è da preferire specialmente quando si lavora su reti reali dove un gran numero di percorsi condividono molti archi (Bovy, 2009; Cascetta, et al., 1997). Un modello di scelta del percorso che ignori l'effetto della tecnica di generazione utilizzata rischia di essere seriamente mis-specificato (Ben Akiva & Boccara, 1995), come provato empiricamente da alcuni autori (Stopher, 1980; Williams & Ortuzar, 1982; Swait & Ben-Akiva, 1987; Başar & Bhat, 2004). Alcuni altri ricercatori hanno inoltre provato che le stime dei parametri del modello sono affette dalla dimensione e composizione dell'insieme di scelta (Bekhor, et al., 2006; Prato & Bekhor, 2007). Ancora, l'insieme di scelta dovrebbe includere tutte le alternative rilevanti, altrimenti le probabilità potrebbero essere affette da errore e quindi non rappresentative (Bliemer & Bovy, 2008).

L'obiettivo di questa applicazione si concentra sull'errore che può essere introdotto nelle stime modellistiche durante il processo di generazione dell'insieme di scelta. Nello specifico, tra tutte le tecniche presenti in letteratura (Prato, 2009), l'attenzione è dedicata alla generazione stocastica e alla correzione per le diverse probabilità di selezione dei percorsi, dovute all'utilizzo di questa tecnica che, inoltre, è molto semplice da implementare anche in caso di reti reali e vaste. La tecnica di generazione stocastica consiste nella ripetuta ricerca del percorso di minimo costo nella rete, il cui calcolo avviene attraverso l'estrazione casuale dei costi sugli archi da apposite funzioni di probabilità. La selezione di queste ultime è una fase cruciale del processo, poiché i parametri con cui sono definite influenzano i percorsi generati, per cui è buona norma identificarli in maniera opportuna in modo da produrre un insieme di scelta plausibile per la stima di modelli su reti a vasta scala. Nel caso esaminato si utilizza una distribuzione gamma che è additiva sugli archi e non consente l'estrazione di costi negativi, per la quale sono stati testati diversi valori dei parametri in modo tale da trovare quelli che offrono i migliori risultati nella riproduzione dei percorsi osservati.

Una volta che viene definita la distribuzione di probabilità, l'insieme di scelta conterrà i percorsi unici attrattivi che sono consistenti con il comportamento osservato. Tuttavia, la generazione stocastica è un problema noto in letteratura come "importance sampling", perché la probabilità di estrazione di un percorso dipende dalle proprietà dello stesso, come i costi. In accordo con quanto argomentato finora, in questa ricerca si propone una metodologia per la derivazione di un fattore di correzione per il campionamento in caso di utilizzo della tecnica di generazione stocastica, per tenere conto delle diverse probabilità di

selezione ed evitare errori nelle stime dei parametri (Frejinger, et al., 2009; Bovy, et al., 2009). L'unico fattore di correzione conosciuto è stato derivato per l'algoritmo di random walk (Frejinger, et al., 2009), ma la sua applicazione su reti vaste appare problematica a causa della natura combinatoria dell'algoritmo, che richiede lunghi tempi computazionali e produce percorsi lunghi e con loop (Kaplan & Prato, 2012; Rieser-Schüssler, et al., 2013). In questa ricerca si propone una metodologia per il calcolo delle probabilità di selezione del percorso quando è utilizzata la generazione stocastica, seguendo le basi della teoria di calcolo del percorso di minimo costo, in modo da poter utilizzare il fattore di correzione nelle stime modellistiche.

Il metodo proposto è applicato sull'intero database di dati a disposizione per la modalità auto guidatore (vedi capitolo 3), con un totale di 3656 spostamenti utilizzabili per le stime modellistiche. L'utilizzo di dati GPS incrementa inoltre la difficoltà nell'applicazione delle tecniche modellistiche alla scelta del percorso, perché l'abbondanza e accuratezza delle informazioni disponibili rende necessario lavorare su reti ad alta risoluzione, con conseguente aumento del tempo computazionale degli algoritmi. Sono stati quindi stimati diversi modelli in modo da identificare i migliori valori per la definizione delle probabilità di selezione del percorso, confrontando le diverse specificazioni in caso di implementazione o meno del fattore di correzione. Il contributo di questo lavoro è quindi quello di: i) incrementare la conoscenza nell'applicazione della tecnica di generazione stocastica, poiché questa è la più efficiente per generare gli insiemi di scelta; ii) proporre una procedura semplice e gestibile dal punto di vista computazionale per il calcolo delle probabilità di selezione del percorso e quindi il fattore di correzione per le stime modellistiche; iii) provare l'efficienza della metodologia proposta su dati di preferenze rilevate in una rete urbana densa.

## **5.1. Contesto teorico**

Come già ampiamente argomentato, la scelta del percorso è tipicamente modellata usando dei modelli di scelta discreta che la rappresentano utilizzando un esaustivo e mutuamente esclusivo insieme di alternative (Ben-Akiva & Lerman, 1985). Il modello più comunemente utilizzato per trattare un numero elevato di alternative è il MNL e le sue modificazioni, così come presentato nel capitolo 2.4.

Teoricamente, l'approccio alla modellazione della scelta del percorso consiste nella soluzione al paradigma di Manski presentato nel capitolo 2.2.5, cui si rimanda per ulteriori approfondimenti. Tra tutte le tecniche di generazione dell'insieme di scelta presenti in letteratura, utili per procedere all'applicazione della metodologia di soluzione del paradigma,

quella stocastica sembra quella che offre i migliori risultati in termini di riproduzione delle osservazioni, semplicità di implementazione e tempi computazionali in caso di applicazione su reti a larga scala (Kaplan & Prato, 2012; Rieser-Schüssler, et al., 2013; Bekhor & Prato, 2009).

Sfortunatamente, la generazione stocastica produce generalmente un grande numero di percorsi e stimare un MNL su un sottoinsieme, che dovrebbe produrre stime consistenti se i percorsi avessero uguali probabilità di campionamento (McFadden, 1978), è inadeguato nel contesto della scelta del percorso (Daganzo & Sheffi, 1977). Ancora, la generazione stocastica produce percorsi con diverse probabilità di campionamento perché, come anticipato nel paragrafo precedente, è un caso di “importance sampling”. Quindi i modelli di scelta del percorso basati su di essa hanno bisogno di un fattore di correzione. L’unica soluzione conosciuta a questo problema è l’algoritmo di random walk (Frejinger, et al., 2009), che calcola le probabilità di estrazione dei costi di percorso su ogni arco della rete. Cominciando dal nodo origine, l’algoritmo sceglie in maniera casuale il nodo successivo secondo una distribuzione di Kumaraswamy, in funzione del peso di ciascun arco in uscita da quel nodo, calcolato come il rapporto tra il costo del percorso di costo minimo tra la coppia O/D e quello più breve che includa al suo interno l’arco considerato. Questa procedura è reiterata fino a quando non si raggiunge il nodo destinazione, e la probabilità di selezione del percorso è data dal prodotto di quelle degli archi che lo compongono. In questo modo è possibile applicare la correzione per il campionamento. Questa metodologia è stata tuttavia testata solo su dati sintetici, mentre le applicazioni su reti a vasta scala hanno mostrato problemi dal punto di vista dei tempi computazionali, i percorsi creati presentano spesso dei circuiti e a volte non si raggiunge il nodo destinazione in un numero ragionevole di iterazioni.

## 5.2. Metodo

Lo studio presentato applica la generazione stocastica dei percorsi (Daganzo & Sheffi, 1977; Nielsen, 2000; Ramming, 2002; Bliemer & Taale, 2006; Bovy & Fiorenzo-Catalano, 2007), approccio simulativo che prevede la ricerca continuativa del percorso di minimo costo attraverso una randomizzazione dei costi sugli archi. L’assunzione comportamentale alla base di questa tecnica è che gli utenti percepiscano i costi del percorso con un errore, rappresentato tramite l’estrazione delle funzioni di costo generalizzato da distribuzioni di probabilità (Prato, 2009). La ricerca del percorso di minimo costo è ripetuta per ogni estrazione degli attributi degli archi sulla rete, aggiungendo l’alternativa all’insieme di scelta solo se non è stata estratta precedentemente. Conseguentemente, l’algoritmo ha la potenzialità di estrarre N percorsi pari al numero di estrazioni casuali, tuttavia lo stesso percorso può essere trovato più volte nel corso delle iterazioni. In questo senso. La

generazione stocastica è un esempio di campionamento con sostituzione, in quanto la rete non viene modificata di iterazione in iterazione per prevenire la selezione di un percorso che è già stato generato. La procedura termina quando viene raggiunto il numero massimo di estrazioni o quando l'insieme di scelta raggiunge le dimensioni desiderate.

Applicare la generazione stocastica consiste inizialmente nel decidere il numero di estrazioni casuali e la distribuzione di probabilità. Riguardo al primo tema, in letteratura sono presenti diversi approcci, che variano tra le 20 e le 200 estrazioni (Kaplan & Prato, 2012; Prato & Bekhor, 2006; Nielsen, 2000; Ramming, 2002; Frejinger & Bierlaire, 2007), mostrando inoltre che l'aumento del numero di iterazioni e, quindi, i tempi computazionali, non necessariamente porta ad un incremento del numero dei percorsi unici generati. In accordo, in questo studio si è deciso di compiere 100 estrazioni casuali in modo tale da contenere lo sforzo computazionale. Riguardo la distribuzione di probabilità, la letteratura suggerisce di utilizzare delle distribuzioni limitate che permettano l'estrazione di costi sugli archi positivi e ragionevoli (Kaplan & Prato, 2012; Prato & Bekhor, 2006; Nielsen, 2000; Ramming, 2002; Frejinger & Bierlaire, 2007). In questo studio si usa quindi una distribuzione gamma parametrizzata da una coppia di numeri positivi  $(k, \theta)$ ,  $k$  è la forma e  $\theta$  la scala. La media è data dal prodotto  $k\theta$ , mentre la varianza è  $k\theta^2$ . Sono state testate diversi valori di parametrizzazione perché una varianza bassa potrebbe generare pochi percorsi unici, mentre valori alti potrebbero portare alla generazione di molti percorsi irrealistici (Prato, 2009). Media e varianza sono state quindi espresse in questo modo:

$$E[X] = k\theta = \text{time} \quad (40)$$

$$\text{Var}[X] = k\theta^2 = (\alpha \cdot \text{time})^2 \quad (41)$$

dove  $\alpha$  è un parametro da cui dipende la varianza della distribuzione e  $\text{time}$  è il tempo di viaggio sull'arco considerando la rete carica.

In questa applicazione sono stati testati diversi valori di  $\alpha$  in modo da avere differenti insiemi di scelta con diverse caratteristiche: 0.25, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, 3.00, 5.00. Sono stati quindi confrontati i percorsi per ogni osservazione con quelli realmente utilizzati: idealmente, l'algoritmo dovrebbe replicare il percorso osservato generandone uno che corrisponda arco per arco; praticamente, l'algoritmo riproduce parte di questo percorso, la cui sovrapposizione viene calcolata come il rapporto tra la lunghezza degli archi comuni e quella totale del percorso osservato. Infine, viene calcolata la "coverage" dell'algoritmo per ciascuno dei valori di parametrizzazione come la percentuale di osservazioni per le quali impostazioni di calcolo permettono un raggiungimento di una soglia di sovrapposizione

prefissata (Bekhor, et al., 2006; Ramming, 2002). Sarà selezionato l'insieme di scelta che meglio riproduce i dati osservati, sulla base della coverage,

Una volta quindi definito il numero di estrazioni e la distribuzione di probabilità, è stato costruito l'insieme di scelta  $CS_i$  per ciascun individuo  $i$ . Considerando  $n$  come l'iterazione corrente,  $N$  come il massimo numero di iterazioni e  $k_{ri}$  come il numero di volte che il percorso  $r$  è stato estratto per l'individuo  $i$ , la procedura è di seguito riassunta:

1. porre  $n=1$  and  $N=100$ ;
2. per ogni arco  $a$  della rete, estrarre il tempo di viaggio  $t^n$  dalla distribuzione gamma;
3. per ogni individuo  $i$ , calcolare il percorso minimo  $r^n_i$  usando l'algoritmo di Dijkstra (1959), sulla base del tempo di viaggio;
4. se  $r^n_i$  non è stato ancora estratto nelle precedenti iterazioni, aggiungerlo a  $CS_i$  e quindi  $k_{ri} = 1$ ; se  $r^n_i$  è già stato estratto, allora  $k_{ri} = k_{ri} + 1$ ;
5. se  $n=N$  andare al punto 6, altrimenti  $n = n+1$  e ripetere dal punto 2;
6. se il percorso estratto non è stato estratto durante le  $N$  iterazioni, aggiungerlo a  $CS_i$ .

La procedura descritta è semplice da implementare poiché richiede solo un generatore di numeri casuali e l'algoritmo di Dijkstra, che permette all'intero processo di essere veloce ed evita la presenza di loop irragionevoli nei percorsi generati.

Un volta terminata questa fase è stata calcolato il fattore di correzione per il campionamento. McFadden (1978) ha mostrato che, usando il MNL su un campione di alternative invece che sull'insieme universale, è possibile ottenere parametri stimati consistenti attraverso la massimizzazione di una quasi-log-verosimiglianza costruita usando le probabilità condizionali, che rende stimatori consistenti se il modello logit ha coefficienti fissi. La probabilità di scelta è espressa quindi dalla seguente formula:

$$P_i(r | CS_i) = \frac{e^{V_{ri} + \ln P(CS_i|r)}}{\sum_{j \in CS_i} e^{V_{ji} + \ln P(CS_i|j)}} \quad (42)$$

dove  $P(CS_i | r)$  è il fattore di correzione per il campionamento.

Una formulazione per il calcolo di quest'ultimo è stata proposta per metodo random walk (Frejinger, et al., 2009), i cui elementi chiave sono le probabilità di selezione del percorso (probabilità di generazione o selezione) e il numero di volte che il percorso è stato estratto (campionamento con sostituzione):

$$P(CS_i | r_i) = \frac{k_{ri}}{q(r_i)} \quad (43)$$

dove  $k_{ri}$  è il numero di volte che il percorso  $r$  appare in  $CS_i$  e  $q(r_i)$  è la probabilità di selezione del percorso.

Mentre  $k_{ri}$  è (già) noto una volta che l'insieme di scelta è stato generato, nella presente ricerca è stata sviluppata una metodologia per calcolare le probabilità di selezione del percorso per i casi in cui si usa la procedura di generazione stocastica. Il principio alla base è lo stesso, ossia l'estrazione del percorso di minimo costo: (i) ad ogni percorso osservato viene associato il tempo di viaggio (basato sugli attributi della rete); (ii) il percorso di minimo costo è anche quello con la maggiore probabilità di selezione, perché la media della distribuzione (corrispondente al tempo di viaggio del percorso più breve) è la più probabile ad essere estratta; (iii) la probabilità di selezione decresce all'aumentare del tempo di viaggio; (iv) i parametri della distribuzione di probabilità associata all'insieme di scelta generato sono calcolati usando come riferimento il percorso più breve. È stata quindi usata una distribuzione gamma parametrizzata dalla stessa coppia  $(k, \theta)$ , in modo da mantenere la consistenza con la tecnica di generazione stocastica. A questo punto rimane da risolvere il problema della continuità della funzione gamma, che ne impedisce il calcolo della probabilità in un singolo punto. Per estrapolare quindi le probabilità di selezione del percorso, è stata utilizzato l'intervallo definito dalla differenza tra la funzione cumulativa associata al valore di riferimento (tempo di viaggio del percorso) incrementata e decrementata, rispettivamente, di un intervallo  $dx$ , funzione della deviazione standard dei tempi di viaggio di tutti i percorsi appartenenti all'insieme di scelta, in modo da approssimare il valore reale nel miglior modo possibile. Sono stati testati differenti dimensioni dell'intervallo:

$$dx = \omega \cdot \sigma_{CS_i} \quad (44)$$

dove  $\omega = (0.1, 0.05, 0.01, 0.005, 0.001, \dots, 10^{-8})$ . La probabilità di un percorso è quindi definita dalla formula seguente:

$$q(r_i) = F(\text{time}_r + dx_{CS_i}) - F(\text{time}_r - dx_{CS_i}) \quad (45)$$

Un altro dei problemi legati alla scelta del percorso è l'elevata correlazione tra le diverse alternative, data la possibilità di sovrapposizione tra i percorsi (Bovy, 2009). Per questo motivo, saranno tenuti in considerazione dei fattori di correzione da introdurre nella funzione di utilità. Si utilizza quindi il termine di correzione Path Size (PS) (Ben-Akiva & Bierlaire, 1999) per ridurre l'utilità di un percorso in funzione del livello di similarità con gli altri del set di scelta a cui appartiene. La formulazione utilizzata è quella originale (Ben-Akiva & Bierlaire, 1999), che è stato dimostrato essere da preferire a quella generalizzata (Prato, 2009) e che inoltre è implementabile nella struttura modellistica logit:

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al}} \quad (46)$$

dove  $\Gamma_k$  è l'insieme degli archi del percorso  $k$ ,  $L_a$  è la lunghezza dell'arco  $a$ ,  $L_k$  è la lunghezza del percorso e  $\delta_{al}$  vale 1 se l'arco  $a$  appartiene al percorso  $l$ . Il path size rappresenta la frazione di percorso che costituisce un'alternativa completa, quindi vale 1 se il percorso è unico, mentre  $N$  percorsi duplicati condivideranno la dimensione  $1/N$ . Data inoltre la disponibilità della probabilità di selezione, è possibile implementare anche una diversa forma della correzione path size. Frejinger et al (2009) hanno proposto una modifica alla formulazione path size originale, assumendo che l'insieme di scelta reale sia quello universale, con la conseguenza che il PS debba riflettere la correlazione tra tutti i percorsi. La versione corretta dell'attributo è chiamata Expanded Path Size (EPS), che tiene conto di un fattore di espansione che corregge per il campionamento:

$$EPS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al} \Phi_l} \quad (47)$$

$\Phi_l$  è il fattore di espansione. La formulazione di quest'ultimo è:

$$\Phi_l = \begin{cases} 1 & \text{se } \delta_{ic} = 1 \text{ o } q(r_i)R_n \geq 1 \\ \frac{1}{q(r_i)R_n} & \text{altrimenti} \end{cases}$$

Dove  $R_n$  è il numero di volte che il percorso è stato estratto, mentre  $\delta_{ic} = 1$  quando  $i$  coincide con il percorso realmente osservato.  $\Phi_l$  vale quindi 1 se il percorso attuale è estratto almeno una volta.

### 5.2.1. Descrizione dei dati

Il metodo proposto è applicato sull'intero database di dati a disposizione per la modalità auto guidatore (vedi capitolo 3), con un totale di 3656 spostamenti utilizzabili per le stime modellistiche. Il campione è eterogeneo, composto dal 50% di maschi e 50% di femmine, dei quali il 36% ha età compresa tra 18-30 anni, il 31% tra 31-40, il 32% tra 41-60 e circa l'1% supera i 61 anni- Tra i partecipanti, il 25% sono lavoratori autonomi, il 51% sono lavoratori dipendenti, il 19% sono studenti mentre il restante 5% è composto da disoccupati. La tabella seguente mostra un riassunto delle caratteristiche del campione utilizzato in relazione alle

variabili che saranno implementate nel modello. Va sottolineato che c'è un'altra categoria stradale (ZTL, zone a traffico limitato) e alcune altre strade di livello inferiore che non sono presenti in questa descrizione ma sono incluse nel modello.

Attribute	Mean	Std.Dev
Travel Time [min]	16.164	14.890
Distance [km]	6.952	8.010
Average V/C ratio	0.526	0.208
% length V/C > 0.8	0.219	0.225
% time dual carriageway	0.235	0.269
% time interdistrict roads	0.391	0.303
% time intradistrict roads	0.218	0.255
% time local roads	0.057	0.160
% time primary suburban roads	0.029	0.122
% time secondary suburban roads	0.063	0.183
Traffic lights per km	0.724	0.854
Give-way signs per km	3.195	2.081
Number of right turns per km	1.019	0.826

Tabella 13 - Caratteristiche dei dati

### 5.2.2. Risultati del modello

Il modello ha girato in circa 1 ore in un personal computer equipaggiato con un processore Intel Q9400 2.66 GHz con 4GB di RAM e un hard disk classico (leggere: non uno allo stato solido). Applicando la metodologia descritta in precedenza e variando i parametri della distribuzione gamma sono stati generati i diversi insiemi di scelta. Per la parametrizzazione delle equazioni 40 e 41 è stato usato il tempo di viaggio calcolato attraverso il modello di macrosimulazione dell'area vasta di Cagliari (vedi capitolo 4.1), che ha permesso inoltre di estrarre le informazioni legate al rapporto volume/capacità (V/C). L'efficacia degli algoritmi con i diversi valori della gamma è definita in termini di coverage dei percorsi osservati, calcolata attraverso il confronto con quelli generati. La Figura 21 mostra l'andamento dell'indicatore, il cui valore più alto (considerando una soglia del 75%) è ottenuto per un  $\alpha = 1.5$ , che sarà poi usato per il proseguo del processo modellistico. La Tabella 14 mostra le stime dei parametri del primo modello (PS-Logit Base) che è un semplice PS-Logit che implementa la formulazione PS originale (Swait & Ben-Akiva, 1987) e non considera la



correzione per il campionamento. Questo sarà il modello da utilizzare come base di confronto per valutare l'effetto delle probabilità di selezione del percorso e del fattore di correzione.

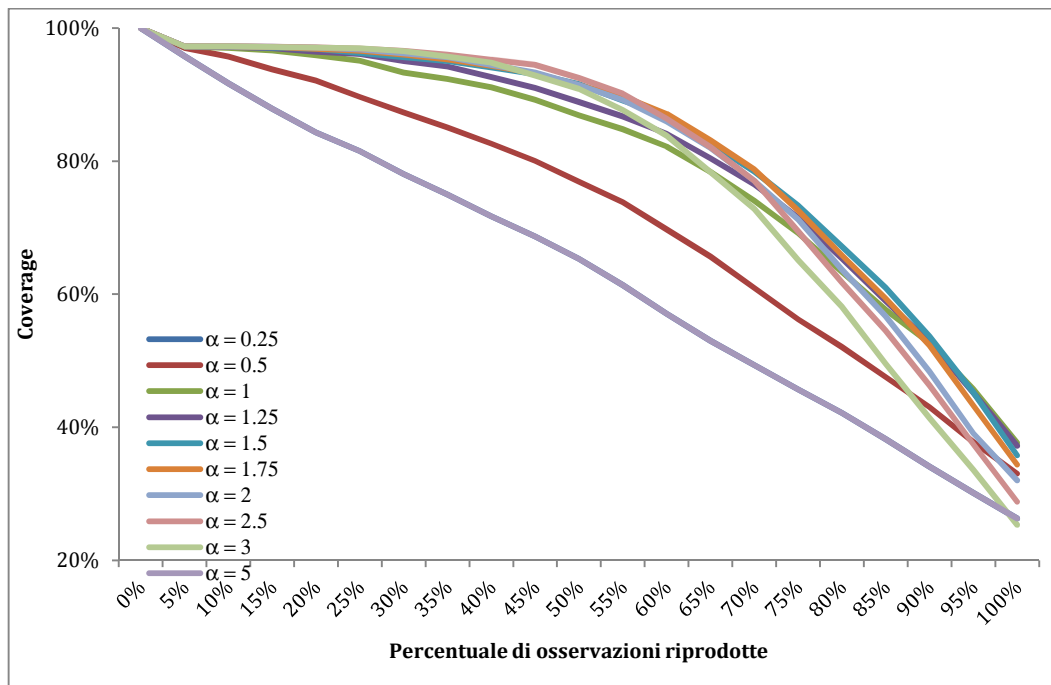


Figura 21 - Coverage degli insiemi di scelta

Variable	est.	std.err.	t-stat.
Travel Time [min]	0.12	0.01	17.70
Average V/C ratio	13.80	0.44	31.80
% length V/C > 0.8	4.20	0.38	10.97
% time dual carriageway	0.70	0.49	1.44
% time interdistrict roads	0.44	0.46	0.96
% time intradistrict roads	-0.45	0.47	-0.96
% time local roads	2.13	0.49	4.36
% time primary suburban roads	1.10	0.78	1.42
% time secondary suburban roads	-0.92	0.60	-1.53
Traffic lights per km	-1.01	0.06	-16.81
Give-way signs per km	0.17	0.02	6.83
Number of right turns per km	-1.29	0.05	-26.68
Number of left turns per km	-1.71	0.05	-32.21
Ln (correction parameter)	-	-	-
Ln (path size)	-0.67	0.04	-15.86
Number of estimated parameters			14
Number of observations			3656
Null log-likelihood			-15703
Final log-likelihood			-8481
Adjusted rho-square			0.459

Tabella 14 - Modello PS-Logit Base senza correzione per il campionamento

I risultati mostrano come le stime dei parametri, per alcune variabili, hanno il segno atteso, come il numero dei semafori per km o il numero di svolte per km, ma per altre importanti variabili come il tempo di viaggio e il path size i segni sono illogici. Oltre alle ragioni teoriche di cui sopra che richiedono la corretta specificazione del modello, questi segni illogici mostrano anche ragioni pratiche che suggeriscono la necessità di verificare se l'applicazione del fattore di correzione proposto potrebbe ridurre o eliminare la distorsione introdotta dalla procedura di generazione stocastica. Si sottolinea il fatto che la distanza non è stata inclusa nelle stime, nonostante fosse stata considerata inizialmente, a causa di diversi problemi di convergenza emersi a causa dell'elevata correlazione con il tempo di viaggio, che risultava in numerose iterazioni e diversi problemi nei parametri stimati. Quindi, la distanza è stata rimossa dalle stime modellistiche finali, che sono arrivate a convergenza una volta risolto questo problema di multi-collinearità.

Per definire la struttura del modello finale, è stato determinato il valore del fattore  $\omega$  per calcolare la probabilità di selezione. Il calcolo è stato eseguito non solo per il fattore di

correzione di campionamento, ma anche per avere l'EPS. Quindi, a partire dal PS-Logit Base, sono stati stimati diversi modelli per differenti valori di  $\omega = (0.1, 0.05, 0.01, 0.005, 0.001, \dots, 10^{-8})$ . Al diminuire di  $\omega$ , l'intervallo dx diventa più piccolo e così come  $q(r_i)$ , che indica che il fattore  $\omega$  influenza la dimensione numerica della probabilità di selezione, per la cui valutazione sono analizzate separatamente due serie di modelli. Per tutto il resto di questo capitolo, si definiscono i modelli PS-Logit che includono la correzione di campionamento come Corrected PS-Logit. Tabella 15 mostra che i modelli che implementano solo la correzione di campionamento non sono influenzati dalla dimensione numerica della probabilità, ottenendo gli stessi parametri di stima finali.

	$\omega = 0.1$	$\omega = 0.05$	$\omega = 0.01$	$\omega = 0.005$	$\omega = 0.001$	(...)	$\omega = 10^{-8}$
Sampling correction test							
Null log-likelihood	-15703	-15703	-15703	-15703	-15703	(...)	-15703
Final log-likelihood	-8174	-8173	-8173	-8173	-8173	(...)	-8173
Rho-square	0.479	0.480	0.480	0.480	0.480	(...)	0.480
Adjusted rho-square	0.479	0.479	0.479	0.479	0.479	(...)	0.479
EPS test							
Null log-likelihood	-15703	-15703	-15703	-15703	-15703	(...)	-15703
Final log-likelihood	-8595	-8586	-8011	-7713	-7159	(...)	-6242
Rho-square	0.453	0.453	0.490	0.509	0.544	(...)	0.602
Adjusted rho-square	0.452	0.452	0.489	0.508	0.543	(...)	0.602

Tabella 15 - Test per la probabilità di selezione

E' chiaro però che i modelli EPS sono sensibili a  $\omega$ . In particolare, questi modelli mostrano un miglioramento degli indicatori di stima al diminuire di  $\omega$ , e così pure la dimensione numerica della probabilità di selezione. Va notato che la bontà di adattamento dei modelli EPS e del Corrected PS-Logit sono migliori rispetto al modello di base. Intuitivamente, il miglioramento aumenta perché al diminuire di  $\omega$  maggiore è la precisione della probabilità. Tuttavia, l'andamento del miglioramento della log-likelihood è logaritmico, con un asintoto orizzontale, e il limite inferiore è stato identificato essere  $\omega=10^{-8}$ . Va inoltre riportato che le stime dei modelli EPS e degli EPS-Logit corretti sono migliori di quello base. Come sommario

finale, sono stati confrontati i risultati delle specificazioni con  $\omega=0.001$  e  $\omega=10^{-8}$ . La Tabella 16 riporta le stime dei parametri del EPS-Logit corretto che implementa sia il EPS sia la correzione di campionamento, dove è evidente che i risultati dei due modelli siano pressoché identici, fatta eccezione per il fattore EPS, che passa da 1.12 a 0.56. I risultati mostrano chiaramente l'efficacia della metodologia proposta. Infatti, vi è un passaggio dal modello di base in cui i segni delle variabili fondamentali come tempo di viaggio e path size sono rispettivamente positivo e negativo, quindi incompatibili con i valori teorici attesi, a questo modello in cui i segni sono coerenti con quelli attesi. Il modello finale mostra anche un notevole miglioramento del valore della verosimiglianza e dell'indice di bontà di adattamento, considerando che è stimato lo stesso numero di parametri.

Model	$\omega = 0.001$			$\omega = 10^{-8}$		
	est.	std.err.	t-stat.	est.	std.err.	t-stat.
Travel Time [min]	-0.18	0.01	-24.18	-0.23	0.01	-23.99
Average V/C ratio	12.30	0.46	26.80	12.30	0.51	24.19
% length V/C > 0.8	2.85	0.40	7.14	3.72	0.44	8.40
% time dual carriageway	-1.96	0.60	-3.26	-1.69	0.77	-2.20
% time interdistrict roads	-0.86	0.57	-1.50	-0.70	0.74	-0.94
% time intradistrict roads	-1.47	0.58	-2.53	-1.44	0.75	-1.92
% time local roads	1.26	0.61	2.08	1.39	0.77	1.81
% time primary suburban roads	0.32	0.85	0.38	0.70	1.01	0.69
% time secondary suburban roads	-2.23	0.70	-3.20	-2.49	0.86	-2.89
Traffic lights per km	-0.51	0.06	-8.23	-0.57	0.07	-8.21
Give-way signs per km	0.26	0.02	10.38	0.16	0.03	5.92
Number of right turns per km	-1.19	0.05	-23.91	-1.21	0.05	-22.00
Number of left turns per km	-1.55	0.05	-29.19	-1.48	0.06	-25.80
Ln (correction parameter)	1.00	fixed	fixed	1.00	fixed	fixed
Ln (EPS)	1.12	0.02	45.54	0.56	0.02	34.40
Ln (path size)	-	-	-	-	-	-
Number of estimated parameters	14			14		
Number of observations	3656			3656		
Null log-likelihood	-15703			-15703		
Final log-likelihood	-6759			-5866		
Adjusted rho-square	0.569			0.626		

Tabella 16 - Stime dei parametri del modello Corrected EPS-Logit

Analizzando le variabili del modello, si nota che gli indicatori relativi alla congestione presentano dei parametri con il segno positivo, mostrando che gli utenti sembrano scegliere percorsi congestionati. Questo potrebbe essere spiegato dal fatto che le persone tendono ad utilizzare i percorsi più familiari che forse sono considerati più sicuri e affidabili rispetto alle alternative disponibili. Inoltre, gli utenti spesso non sono a conoscenza di tutte le opzioni disponibili, che necessariamente portano a scelte più "standard". I segni dei parametri delle variabili associate al tempo di viaggio sono negativi, con l'eccezione della percentuale dedicata alle strade extraurbane principali e strade locali. Una possibile spiegazione potrebbe essere che, come illustrato nella descrizione dei dati, la rete stradale dell'area metropolitana di Cagliari è costituita in gran parte da strade locali, quindi è molto probabile che queste costituiscano una parte importante dei percorsi utilizzati dagli utenti, da cui il segno positivo del parametro associato. Inoltre, poiché i dati osservati si riferiscono a spostamenti urbani, è possibile che le strade principali siano a volte poco utilizzate, come mostrato anche dall'analisi dei dati. Lo stesso vale per il numero di segnali di precedenza per km, mentre il numero di semafori per km, con segno negativo, mostra che gli utenti percepiscono questi come un ostacolo e tendono ad evitarli. Una considerazione interessante nasce invece per i valori delle variabili associate alle svolte, che infatti mostrano (come previsto) che vi è una tendenza a evitare svolte per quanto possibile, confermando la teoria che gli utenti preferiscono percorsi rettilinei (Conroy Dalton, 2003; Raghbir & Krishna, 1996). È inoltre chiaro che le svolte a sinistra producano una maggiore disutilità, spiegato con il fatto che, in fondo, girare a sinistra è più pericoloso che a destra, perché spesso implica l'interazione con il traffico in arrivo dal senso opposto.

## 6. CONCLUSIONI

Analizzare e comprendere i diversi aspetti legati alla scelta del percorso, come ampiamente descritto, è fondamentale per essere in grado di stare al passo dell'attuale sviluppo tecnologico (ITS, AMTS, etc.), che ha messo a disposizione informazioni sempre più precise ma che, a sua volta, ha l'esigenza di essere alimentato da altrettanto precisi dati su come gli utenti si comportino e prendano le proprie decisioni. Le nuove tecniche di acquisizione dati offrono un notevole supporto in questa direzione, tuttavia necessitano di essere accompagnate da tecniche di analisi e modellazione avanzate per ottenere i riferimenti utili ad essere implementati in questi contesti operativi. La presente tesi di Dottorato di Ricerca si pone quindi l'obiettivo di offrire un contributo al tema, sviluppando una procedura in grado di riprodurre la scelta del percorso e che offra allo stesso tempo garanzie di innovazione e praticità. Da un lato è proposto l'utilizzo di una base dati acquisita tramite un sistema GPS di ultima generazione, le cui caratteristiche (qualità, quantità e precisione) fanno sì che anche il contesto sul quale si lavora (rete di trasporto) sia ad alta risoluzione, esteso e rappresentativo della realtà. Dall'altro lato viene sviluppata una metodologia modellistica che, sulla base dello stato attuale della ricerca sulla scelta del percorso, ottimizzi l'utilizzo di questa tipologia di dati ad alta precisione, colmando le lacune presenti nelle specifiche modellistiche attuali e arrivando alla definizione di una procedura operativa efficace e facilmente applicabile in contesti operativi pratici.

La struttura del lavoro è stata quindi organizzata per rispondere alle esigenze derivate dagli obiettivi prefissati. A partire da una riflessione sul fenomeno della scelta del percorso secondo tutte le sue sfaccettature, presentando le diverse teorie sviluppate finora e i risultati ottenuti, segue un approfondimento circa la raccolta dei dati, riportando le caratteristiche delle tecniche di analisi usate in passato e andando a descrivere nel dettaglio le potenzialità offerte dai dispositivi GPS. Nelle ricerche precedenti l'argomento maggiormente trattato ha riguardato il confronto tra i percorsi osservati e quelli di costo minimo, calcolati utilizzando piattaforme GIS attraverso la minimizzazione di alcuni attributi (come tempi e distanze di percorrenza) che vengono comunemente utilizzati nelle procedure di modellazione dei flussi di traffico (assegnazione). Utilizzando il database di spostamenti ottenuto attraverso la campagna d'indagine "Casteddu Mobility Styles" (vedi cap.3), nella prima fase di questo lavoro di ricerca si è deciso di replicare questi confronti, con l'obiettivo di verificare le caratteristiche dei dati di CMS ed effettuare una comparazione con le ricerche esistenti. Il limite principale dei precedenti lavori è dato dall'utilizzo di piattaforme GIS, che offrono dei risultati generali e non totalmente rappresentativi delle effettive condizioni operative della rete di trasporto. Allo scopo di ovviare a questo problema, si è deciso di utilizzare, per il calcolo dei percorsi simulati, un modello di macrosimulazione del traffico correttamente

calibrato per la rete di trasporto di Cagliari. In questo modo, la base dati simulata è costruita in un contesto operativo simile a quello nel quale sono state raccolte le informazioni, offrendo migliori garanzie di rappresentazione e affidabilità. Per la specifica applicazione sono stati selezionati, dall'intero campione di CMS, 626 spostamenti avvenuti nell'ora di punta del mattino (congruente con l'orario di picco per cui il modello è calibrato). Le analisi hanno mostrato che il modello di simulazione riproduce circa il 78% dei percorsi osservati, in linea con quanto riportato in letteratura da Ortuzar e Willumsen (2001), i quali indicano che le funzioni di costo basate su tempo e distanza riescono a spiegare circa il 60%-80% dei percorsi osservati nella pratica. È stato rilevato come l'attributo che tende ad essere minimizzato maggiormente è la distanza, segno che la percezione spaziale sia particolarmente sviluppata, mentre questo non avviene per quanto riguarda i tempi di percorrenza. Per percorsi abitudinari di tipo casa-lavoro per i quali si possiede una buona conoscenza della rete e delle condizioni del traffico, è stato evidenziato come gli utenti effettuino delle scelte che si avvicinano maggiormente a quelle cosiddette razionali, con dei valori degli indici di deviazione inferiori rispetto a quelli relativi agli spostamenti per attività discrezionali. È stato inoltre registrato che, nei casi di spostamenti ripetuti tra la stessa coppia OD in cui vengono scelti più percorsi, quelli maggiormente utilizzati (con caratteristiche simili a quelli di minimo costo calcolati con il modello di simulazione) abbiano avuto inizio in istanti successivi rispetto a quelli "second best", sperimentando quindi delle condizioni di rete più favorevoli. È interessante comparare le risultanze tra l'indagine pilota effettuata precedentemente da Spissu, et al. (2011) e lo studio attuale: infatti, era stato trovato che, nel caso di individui che ripetono lo stesso percorso tra la medesima coppia OD, il 92,7% ne usasse solo uno e i restanti due o più, mentre questo studio ha mostrato che circa il 29% degli utenti selezionati utilizzi più di un percorso per lo spostamento tra la stessa coppia origine-destinazione; la differenza con l'indagine attuale è quindi di circa il 23% e può essere attribuita al fatto che siano stati studiati esclusivamente i percorsi avvenuti nell'intervallo orario del mattino, per un campione di utenti eterogeneo, mentre nella precedente ricerca il dato è riferito all'intera giornata. Il valore registrato attualmente è simile al 26% trovato da Mannering & Kim (1994), i cui dati sono stati però raccolti attraverso dei questionari, mentre Jan, et al. (2000) mostrano che il 34% usa lo stesso percorso, simile al 40% trovato da Li, et al. (2005) che coincide anche con quanto emerso dagli studi di Zhu & Levinson (2010) e Levinson & Zhu (2013). Questa particolarità di utilizzare più percorsi per spostamenti tra la stessa coppia OD è una caratteristica nota come intravariabilità. Nonostante, come riportato in precedenza, diversi studi con dati GPS abbiano messo in evidenza il fenomeno, nessuno di essi ha approfondito l'argomento. In letteratura esistono altre ricerche che si sono dedicate al tema con un approccio volto alla modellazione dell'intravariabilità (o route switching), ma l'attenzione era principalmente

focalizzata sull'effetto della somministrazione di informazioni di percorso e i dati utilizzati erano raccolti attraverso questionari, interviste o esperimenti in laboratorio. In questa tesi invece si è analizzato il fenomeno dell'intravariabilità usando i dati GPS e applicando un modello di scelta discreta. Il database di spostamenti utilizzato è relativo ad un sottocampione delle 626 osservazioni analizzate in precedenza, relative a 66 utenti per un totale di 361 spostamenti aventi la caratteristica di ripetitività, con l'obiettivo di individuare quali siano le caratteristiche dei percorsi, dello spostamento e socio-economiche degli individui che determinano il fatto di scegliere un percorso abitualmente o meno. Sono stati quindi stimati dei modelli di tipo mixed logit, dai quali è risultato come il numero di semafori per km (negativamente), i tratti di strade a scorrimento veloce (positivamente, così come trovato anche da Zhu e Levinson (2010)) e l'appartenenza del percorso al set di minimo costo (positivamente) influiscano sul fatto che un percorso sia scelto abitualmente. Il livello di congestione, per quanto non significativo e non inserito nella specificazione finale, mostra invece un'influenza positiva, in contrasto ad esempio da quanto riscontrato da Khattak et al (1995), ma in accordo con i risultati di Srinivasan e Mahmassani (2003), che giustificano il fatto con la difficoltà di re-routing nella rete. Tra le variabili socio economiche, è stato trovato che i maschi siano maggiormente propensi a cambiare percorso, risultato in linea con quanto evidenziato da più autori in letteratura (Khattak et al. 1995, Mannering & Kim 1994, Han et al. 2001, Xu et al. 2010). È interessante inoltre riportare come la variazione di percezione dei minuti di km di percorrenza assuma particolare rilevanza statistica, ulteriore conferma di come le differenze di percezione tra gli utenti siano particolarmente significative sui comportamenti di scelta. Inoltre, è stato trovato come l'esperienza alla guida porti ad una maggiore propensione al cambio di percorso, segno che la maggiore conoscenza della rete influisca poi sulle scelte effettuate anche per spostamenti abitudinari e, quindi tendenzialmente meno soggetti a cambiamenti.

Una volta terminate le analisi sui dati GPS così come descritto in precedenza, necessarie per inquadrare le caratteristiche del campione di dati a disposizione, l'obiettivo primario della tesi era quello di stimare un modello di scelta del percorso. Il fenomeno è comunemente rappresentato attraverso un processo a due stadi: generazione dell'insieme di scelta e scelta del percorso. Il primo stadio è particolarmente importante specie in caso di utilizzo di dati ad alta risoluzione e reti reali, poiché necessità di essere rappresentativo e deve necessariamente contenere tutte le alternative rilevanti al fine di evitare il rischio di avere un modello mis-specificato. Nel presente lavoro di ricerca si è posta quindi l'attenzione sull'errore che potrebbe introdursi nelle stime dei modelli in seguito al processo di generazione dell'insieme di scelta, focalizzandosi sulla tecnica di generazione stocastica. Quest'ultima, è una tecnica che è stato mostrato essere particolarmente efficiente nelle applicazioni della scelta del percorso (Kaplan & Prato, 2012; Prato & Bekhor, 2006; Nielsen,



2000; Ramming, 2002; Frejinger & Bierlaire, 2007). Tuttavia, essa è un caso di “importance sampling” dove la selezione del percorso dipende dalle proprietà dello stesso, ragione per cui i modelli di scelta del percorso basati sulla generazione stocastica dovrebbero includere un fattore di correzione che tenga conto delle differenti probabilità di selezione. In questo studio è proposta una metodologia per il calcolo e la considerazione di tale correzione in modelli basati sulla struttura del MNL con insiemi di scelta generati con la tecnica stocastica. Nello specifico, l’attenzione è stata concentrata sul fattore di correzione sviluppato per l’algoritmo di random walk (Frejinger, et al., 2009), con l’obiettivo di poterlo utilizzare anche con la generazione stocastica. Allo scopo, occorre quindi avere a disposizione le informazioni con cui è definito: numero di estrazioni dello stesso percorso e probabilità di selezione dello stesso. Il primo dato è facilmente ricavabile, mentre allo stato attuale non esiste una metodologia per il calcolo delle probabilità di selezione nel caso di applicazione della generazione stocastica. Nella presente ricerca è stata proposta quindi una metodologia per il calcolo di queste ultime, utilizzando come riferimento la teoria del percorso di costo minimo, che è inoltre una componente principale della struttura teorica alla base della stessa tecnica di generazione. Una volta calcolate le probabilità di selezione, è stato possibile ricavare gli EPS (?) e implementare il fattore di correzione nelle funzioni di utilità per la stima dei modelli. La metodologia proposta è stata applicata all’intero database di spostamenti auto-guidatore di CMS, con un totale di 3656 osservazioni utili. I risultati hanno mostrato come la dimensione numerica delle probabilità di selezione del percorso influenzino i modelli EPS, con un aumento della bontà di adattamento al decrescere dell’intervallo su cui sono calcolate le probabilità. Il fattore di correzione in sé, infatti, non è affetto dalla dimensione numerica della probabilità, contribuendo comunque al miglioramento della bontà di adattamento del modello. La tesi quindi contribuisce in maniera molteplice allo stato dell’arte della modellazione della scelta del percorso, fornendo nuovi spunti per l’applicazione della tecnica di generazione stocastica, specialmente su reti reali e di grandi dimensioni dove l’unica necessità è disporre di un generatore di numeri casuale e l’algoritmo di Dijkstra; si propone inoltre una procedura semplice e intuitiva dal punto di vista computazionale, e si prova l’efficienza della stessa per l’applicazione a dati RP in reti urbane dense, mostrando un incremento della bontà di adattamento del modello che inoltre passa da stime dove alcuni parametri fondamentali come il tempo di viaggio hanno segni illogici, ad una specificazione dove tutte le variabili hanno il segno atteso.

Rimangono ancora diversi campi su cui poter approfondire i temi di ricerca, che saranno oggetto di sviluppo futuro. Il confronto dei percorsi di costo minimo, infatti, potrebbe essere effettuato con un modello di simulazione stocastico invece che deterministico, individuando quali siano i parametri sui cui poter agire per migliorarne la rappresentazione. L’analisi dell’intravariabilità degli spostamenti offre invece numerosi spunti su cui riflettere.

Utilizzando l'intero campione di spostamenti, infatti, sarebbe interessante verificare l'effetto dell'affidabilità del tempo di viaggio, fattore che si è dimostrato particolarmente importante nelle analisi comportamentali. Altri fattori, come l'inerzia e l'abitudine, potrebbero inoltre essere derivati e specificati attraverso delle particolari applicazioni modellistiche. Inoltre, un altro effetto da tenere in considerazione è quello legato all'apprendimento, valutando in che modo le scelte effettuate in precedenza influenzino quelle attuali, valutandone la correlazione.

Sarà inoltre interessante testare l'effetto del fattore di correzione su modelli di tipo Mixed Logit, che è stato dimostrato possibile stimare attraverso la massimizzazione di una funzione di log-verosimiglianza modificata (Guevara & Ben-Akiva, 2013), e confrontare i modelli stimati secondo le specificazioni presentate con altri di tipo il Recursive Logit (Fosgerau, et al., 2013), che è un approccio basato sugli archi senza nessun vincolo legato agli insiemi di scelta.

## 7. BIBLIOGRAFIA

- Abdel-Aty, M. et al., 1994. Models of commuters' information use and route choice: initial results based on southern California commuter route choice survey. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1453.
- Akamatsu, T., 1996. Cyclic flows, Markov process and stochastic traffic assignment. *Transportation Research Part B: Methodological*, 30(5), pp. 369-386.
- Akgün, V., Erkut, E. & Batta, R., 2000. On finding dissimilar paths. *European Journal of Operational Research*, 121(2), pp. 232-246.
- Albert, G., Toledo, T. & Ben-Zion, U., 2011. The role of personality factors in repeated route choice behavior: behavioral economics perspective. *European Transport \ Trasporti Europei*, Volume 48, pp. 47-59.
- Arifin, Z. N. & Axhausen, K. W., 2011. Investigating commute mode and route choice variability in Jakarta using multi-day GPS Data. *Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme*.
- Asakura, Y. & Iryo, T., 2007. Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice*, p. 684-690.
- Avineri, E. & Bovy, P. H., 2008. Identification of Parameters for Prospect Theory Model for Travel Choice Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2082, pp. 141-147.
- Avineri, E. & Prashker, J. N., 2003. Sensitivity to Uncertainty: The Need for a Paradigm Shift. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1854, pp. 90-98.
- Azevedo, J., Costa, M., Madeira, J. & Martins, E., 1993. An algorithm for the ranking of shortest paths. *European Journal of Operational Research*, 69(1993), pp. 97-106.
- Bamberg, S., Ajzen, I. & Schmidt, P., 2003. Choice of Travel Mode in the Theory of Planned Behavior: The Roles of Past Behavior, Habit, and Reasoned Action. *Basic and Applied Social Psychology*, 25(3), pp. 175-187.
- Barbeau, S. et al., 2009. TRAC-IT – A Software Architecture Supporting Simultaneous Travel Behavior Data Collection & Real-time Location-Based Services for GPS-Enabled Mobile Phones. *Paper presented at the 88th Annual Meeting of the Transportation Research*, January.

- Başar, G. & Bhat, C., 2004. A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. *Transportation Research Part B: Methodological*, 38(10), p. 889–904.
- Beckx, C., Panis, L., Janssens, D. & Wets, G., 2010. Applying activity-travel data for the assessment of vehicle exhaust emissions: Application of a GPS-enhanced data collection tool. *Transportation Research Part D: Transport and Environment*, 15(2), p. 117–122.
- Bekhor, S., Ben-Akiva, M. & Ramming, S., 2002. Adaptation of logit kernel to route choice situation. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1805, pp. 78-85.
- Bekhor, S., Ben-Akiva, M. & Ramming, S., 2006. Evaluation of choice set generation algorithms for route choice models. *Annals of operation research*, 144(1), pp. 235-247.
- Bekhor, S. & Prashker, J., 2001. Stochastic user equilibrium formulation for the generalized nested logit model. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1752, pp. 84-90.
- Bekhor, S. & Prato, C., 2006. *Effects of choice set composition in route choice modeling*. Kyoto, Japan, Proceedings of the 11th International Conference on Travel Behavior Research.
- Bekhor, S. & Prato, C., 2009. Methodological transferability in route choice modeling. *Transportation Research Part B*, 43(4), pp. 422-437.
- Bekhor, S. & Toledo, T., 2005. Investigating path-based solution algorithms to the stochastic user equilibrium problem. *Transportation Research Part B*, 39(3), pp. 279-295.
- Bellman, R., 1958. On a routing problem. *Quarterly of Applied Mathematics*, 16(1), pp. 87-90.
- Bellman, R. & Kalaba, R., 1960. On k-th best policies. *Journal of the Society for Industrial and Applied Mathematics*, Volume 8, pp. 582-585.
- Bell, M. G., 1995. Alternatives to Dial's logit assignment algorithm. *Transportation Research Part B: Methodological*, 29(4), pp. 287-295.
- Ben Akiva, M. & Bolduc, D., 1996. *Multinomial probit with a logit kernel and a general parametric specification of the covariance structure*, s.l.: s.n.
- Ben Akiva, M. E. & Boccara, B., 1995. Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1), pp. 9-24.
- Ben-Akiva, M., Bergman, M., Daly, A. & Ramaswamy, R., 1984. Modeling inter urban route choice behaviour. *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, VNU Science Press, pp. 299-330.

- Ben-Akiva, M. & Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: *Handbook of Transportation Science*. Kuwer, Dordrecht, The Netherlands: Hall, R.W., pp. 5-34.
- Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. & Mishalani, R., 1998. *DynaMIT: a simulation-based system for traffic prediction and guidance generation*. San Juan, Porto Rico, Paper presented at TRISTAN III.
- Ben-Akiva, M., Bolduc, D. & Daly, A., 1993. Estimation of travel choice models with randomly distributed values of time. *Transportation Research Record: Journal of the Transportation research Board*, Volume 1413, pp. 88-97.
- Ben-Akiva, M., Koutsopoulos, H., Mishalani, R. & Yang, Q., 1997. Simulation Laboratory for Evaluating Dynamic Traffic Management Systems. *ASCE Journal of Transportation Engineering*, 123(4).
- Ben-Akiva, M. & Lerman, S., 1985. *Discrete choice analysis: theory and application to travel demand*. Cambridge, Massachusetts: The MIT Press.
- Ben-Elia, E. & Shiftan, Y., 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A*, Volume 44, p. 249–264.
- Benjamins, M., Lindveld, K. & Van Nes, R., 2002. *Multimodal travel choice modelling: a supernetwork approach*. Washington, D.C., Proceedings of the 81th Annual Meeting of the Transportation Research Board.
- Bierlaire, M. & Frejinger, E., 2005. *Route choice models with subpath components*. Ascona, Switzerland, Proceedings of the 5th Swiss Transport Research Conference.
- Bliemer, M. C. J. & Bovy, P. H. L., 2008. Impact of Route Choice Set on Route Choice Probabilities. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2076, pp. 10-19.
- Bliemer, M. & Taale, H., 2006. *Route generation and dynamic traffic assignment for large networks*. Leeds, UK, Proceedings of the First International Symposium on Dynamic Traffic Assignment, pp. 90-99.
- Block, H. & Marschak, J., 1960. Random orderings and stochastic probabilities of response. In: *Contributions to probability and statistics*. Stanford, CA: Stanford University Press, pp. 97-132.
- Bogers, E. A. I., Viti, F. & Hoogendoorn, S. P., 2005. Joint Modeling of Advanced Travel Information Service, Habit, and Learning Impacts on Route Choice by Laboratory

- Simulator Experiments. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1926, pp. 189-197.
- Bovy, P., Bekhor, S. & Prato, C., 2008. The factor of revised path size: an alternative derivation. *Transportation Research record: Journal of the Transportation Research Board*, Volume 2076, pp. 132-140.
- Bovy, P., Bekhor, S. & Prato, C., 2009. *Route sampling correction for stochastic route choice set generation*. Washington, D.C., Proceedings of the 88th Annual Meeting of the Transportation Research Board.
- Bovy, P. H. & Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica*, 3(3), pp. 173-189.
- Bovy, P. H. L., 2009. On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews*, 29(1), pp. 43-68.
- Bovy, P. H. L. & Stern, E., 1990. *Route Choice: Way Finding in Transport Networks*. Dordrecht, The Netherlands: Kluwer Academic.
- Bricka, S., Zmud, J., Wolf, J. & Freedman, J., 2009. Household Travel Surveys with GPS: An Experiment. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2105, pp. 51-56.
- Burrell, J., 1968. *Multiple route assignment and its application to capacity restraint*. Karlsruhe, Leutzbach, W. and Baron, P., Fourth International Symposium on the Theory of Traffic Flow.
- Camerer, C. F. & Fehr, E., 2006. When does "economic man" dominate social behavior?. *Science*, 311(5757), pp. 47-52.
- Carey, M., 1992. Non convexity of the dynamic assignment problem. *Transportation Research B*, 26(2), pp. 127-133.
- Carrion, C. & Levinson, D., 2012. Valuation of travel time reliability from a GPS-based experimental design. *Transportation Research Part C*.
- Cascetta, E., 2001. *Transportation Systems Engineering: Theory and Methods*. s.l.:Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Cascetta, E., Nuzzolo, A., Russo, F. & Vitetta, A., 1996. *A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks*. Pergamon, Proceedings of the 13th International Symposium on Transportation and Traffic Theory, pp. 697-711.

- Cascetta, E. & Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, 9(4), pp. 249-263.
- Cascetta, E., Russo, F. & Vitetta, A., 1997. *Stochastic user equilibrium assignment with explicit path enumeration: comparison of models and algorithms*. Oxford, U. K, Proceedings of 8th International Symposium on Transportation Systems, pp. 1031-1037.
- Caussade, S., Ortuzar, J., Rizzi, L. & Hensher, D., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*, 39(7), pp. 621-640.
- Christopher, W. E., Bernstein, D. & Kornhauser, A. L., 2000. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8(1-6), pp. 91-108.
- Chu, C., 1989. *A paired combinatorial logit model for travel demand analysis*. Ventura, USA, Proceedings of the 5th World Conference on Transportation Research, pp. 295-309.
- Conroy Dalton, R., 2003. The Secret Is To Follow Your Nose Route Path Selection and Angularity. *Environment and Behavior*, 35(1), pp. 107-131.
- Corona, C., Sanjust, B., Simbula, M. & Trogu, D., 2012. On Implementing a two-step map matching process of Gps-based data on a GIS platform. *Paper presented at the 7th International Conference on Informatics and Urban and Regional Planning*.
- Daganzo, C., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with hydrodynamic theory. *Transportation Research B*, Volume 28, p. 269-287.
- Daganzo, C. F. & Sheffi, Y., 1977. On stochastic models of traffic assignment. *Transportation Science*, Volume 11, pp. 253-274.
- De la Barra, T., Perez, B. & Anez, J., 1993. *Multidimensional path search and assignment*. University of Manchester, United Kingdom, 21st PTRC Summer Annual Meeting.
- Dial, R., 1971. A probabilistic multi-path traffic assignment algorithm which obviates path enumeration. *Transportation Research*, Volume 5, pp. 83-111.
- Dial, R., 1997. Bicriterion traffic assignment: efficient algorithms plus examples. *Transportation Research Part B*, 31(5), pp. 357-379.
- Dial, R. B., 2001. Equilibrium logit traffic assignment: elementary theory and algorithms. *Transportation Research Part B: Methodological*.

- Dial, R. B., 2006. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B: Methodological*, 40(10), pp. 917-936.
- Dijkstra, E., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik*, Volume 1, pp. 267-271.
- Domencich, T. & McFadden, D., 1975. *Urban Travel Demand: a Behavioural Analysis*. Amsterdam: North-Holland.
- Dreyfus, S., 1969. An appraisal of some shortest path algorithms. *Operational Research*, Volume 17, pp. 395-412.
- Einhorn, H. J., 1970. The use of nonlinear, non-compensatory models in decision making. *Psychological Bulletin*, 73(3), pp. 221-230.
- Eppstein, D., 1998. Finding the K shortest paths. *Journal of the Society for Industrial and Applied Mathematics*, 28(2), pp. 652-673.
- Erev, I., Bereby-Meyer, Y. & Roth, A., 1999. The Effect of Adding a Constant to all Payoffs: Experimental Investigation and Implications for Reinforcement Learning Models. *Journal of Economic Behavior and Organization*, Volume 39, pp. 111-128.
- Fafieanie, M., 2009. Calibrating route set generation by map matching GPS data. *Master Thesis, University of Twente, Enschede - The Netherlands*.
- Fiorenzo-Catalano, S. & Van der Zij, N., 2001. *A forecasting model for inland navigation based on route enumeration*. London, Proceedings of the European Transport Conference, PTRC Education and Research Services Ltd, pp. 1-11.
- Fisk, C., 1980. Some developments in equilibrium traffic assignment. *Transportation Research*, Volume 14B, pp. 243-255.
- Fosgerau, M., Frejinger, E. & Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B*, Volume 56, pp. 70-80.
- Frejinger, E., 2007. *Random sampling of alternatives in a route choice context*. Leeuwenhorst, The Netherlands., Proceedings of the European Transport Conference.
- Frejinger, E. & Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B*, 41(3), pp. 363-378.
- Frejinger, E., Bierlaire, M. & M., B.-A., 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B*, Volume 43, pp. 984-994.



- Friedrich, M., Hofsäß, I. & Wekeck, S., 2001. Timetable-based transit assignment using branch and bound techniques. *Transportation Research Record: Journal of the Transportation Research Board*, 1752(1), pp. 100-107.
- Friesz, T. et al., 1993. A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research*, Volume 41, pp. 179-191.
- Fudenberg, D. & Levine, D., 1998. *The Theory of Learning in Games*. MIT press.
- Gliebe, J., Koppelman, F. & Ziliaskopoulos, A., 1999. *Route choice using a paired combinatorial logit model*. Paper presented at the 78th meeting of the Transportation Research Board. Washington DC, Paper presented at the 78th meeting of the Transportation Research Board.
- Guevara, C. A. & Ben-Akiva, M. E., 2013. Sampling of alternatives in Logit Mixture models. *Transportation Research Part B: Methodological*, Volume 58, p. 185-198.
- Hadjiconstantinou, E. & Christofides, N., 1999. An efficient implementation of an algorithm for finding k-shortest simple paths. *Networks*, Volume 34, pp. 88-101.
- Han, B., Algers, S. & Engelson, L., 2001. Accommodating Drivers' Taste Variation and Repeated Choice Correlation in Route Choice Modeling by Using the Mixed Logit Model. *paper presented at the 80th Annual Meeting of the Transportation Research Board, Washington, DC*.
- Hato, E., Shinji, I. & Mitani, T., 2006. Development of MoALs (Mobile Activity Loggers supported by gps-phones) for travel behavior analysis. *Paper presented at the 85th Annual Meeting of the Transportation Research*.
- Hauser, J. R., Ding, M. & Gaskin, S. P., 2009. *Non-compensatory (and compensatory) models of consideration-set decisions*. Delray Beach, FL, Proceedings of the Sawtooth Software Conference.
- Hess, S., Bierlaire, M. & Polak, J., 2005. Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A*, 39(2-3), pp. 221-236.
- Hoffman, W. & Pavley, R., 1959. A method for the solution of the n-th best path problem. *Journal of the Association for Computing Machinery*, Volume 6, pp. 506-514.
- Hoogendoorn-Lanser, S., 2005. *Modelling travel behaviour in multi-modal networks*. Ph.D. Thesis a cura di TRAIL Research School, Technical University of Delft: Delft, The Netherlands.
- Horowitz, J. L., 1984. The stability of stochastic equilibrium in a two-link transportation network. *Transportation Research Part B: Methodological*, 18(1), pp. 13-28.

- Horowitz, J. & Louviere, J., 1995. What is the role of consideration sets in choice modeling?. *International Journal of Research in Marketing*, Volume 12, pp. 39-54.
- Hunt, D. & Kornhauser, A., 1997. Assigning traffic over essentially-least-cost paths.. *Transportation Research Record: Journal of the Transportation research Board*, Volume 1556, pp. 1-7.
- Jan, O., Horowitz, A. & Peng, Z.-R., 2000. Using GPS data to understand variations in path choice. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1725, pp. 37-44.
- Jou, R., 2001. Modeling the impact of pre-trip information on commuter departure time and route choice. *Transportation Research Part B*, 35(10), pp. 887-902.
- Jou, R.-C., 2004. Real time traffic information and its impacts on route switching behavior of expressway drivers. *Journal of advanced transportation*, 38(2), pp. 187-223.
- Kahneman, D. & Tversky, A., 1979. Prospect Theory: An Analysis of Decisions under Risk. *Econometrica*, 47(2), pp. 263-291.
- Kaplan, S. & Prato, C. G., 2012. Closing the gap between behavior and models in routechoice: The role of spatiotemporalconstraints and latent traits in choice set formation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(1), p. 9-24.
- Kaplan, S. & Prato, C. G., 2012. Closing the gap between behavior and models in routechoice: The role of spatiotemporalconstraints and latenttraits in choicessetformation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(1), p. 9-24.
- Katoh, N., Ibaraki, T. & Mine, H., 1982. An efficient algorithm for k shortest simple paths. *Networks*, Volume 12, pp. 411-427.
- Katsikopoulos, K., Duse Anthony, Y., Fisher, D. & Duffy, S., 2002. Risk attitude reversals in drivers' route choice when range of travel time information is provided. *Human factors*, 3(44).
- Khattak, A. J., Schofer, J. L. & Koppelman, F. S., 1991. Effect of Traffic Reports on Commuters' Route and Departure Time Changes. *Proceedings - Society of Automotive Engineers*, pp. 669-679.
- Khattak, A. J., Schofer, J. L. & Koppelman, F. S., 1995. Effect of traffic information on commuters' propensity to change route and departure time. *Journal of Advanced Transportation*, 29(2), p. 193-212.
- Koelman, F. & Wen, C., 1998. Alternative nested logit models: structure, properties and estimation. *Transportation Research Part B*, 32(5), pp. 289-298.

- Kuby, M., Zhongyi, X. & Xiaodong, X., 1997. A minimax method for finding the k-best differentiated paths. *Geographical Analysis*, 29(4), pp. 298-313.
- Lam, T., 2000. Route and Scheduling Choice Under Travel Time Uncertainty. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1725, pp. 71-78.
- Lam, T. & Small, K., 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E*, 37(2-3), pp. 231-251.
- Lawler, E., 1976. *Combinatorial Optimization: Networks and Matroids*. New York: Holt, Rinehart & Winston.
- Leonard, D., Gower, P. & Taylor, N., 1989. CONTRAM: Structure of the model. *Crowthorne, TRRL Research Report*, Volume 178.
- Levinson, D. & Zhu, S., 2011. s.l.:s.n.
- Levinson, D. & Zhu, S., 2013. A Portfolio Theory of Route Choice. *Transportation Research Part C: Emerging Technologies*.
- Lighthill, M. & Whitham, G., 1955. *On kinematic waves II: A theory of traffic flow on long crowded roads*. s.l., . Proceedings of the Royal Society of London, p. 317-345.
- Li, H., Guensler, R. & Ogle, J., 2005. An Analysis of Morning Commute Route Choice Patterns Using GPS Based Vehicle Activity Data. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1926, pp. 162-170.
- Lombard, K. & Church, R., 1993. The gateway shortest path problem: generating alternative routes for a corridor location problem. *Geographical Systems*, Volume 1, pp. 25-45.
- Luce, D., 1959. *Individual choice behavior: a theoretical analysis*. s.l.:Wiley, New York, NY.
- Maher, M., 1998. Algorithms for logit-based stochastic user equilibrium assignment. *Transportation Research Part B: Methodological*, , 32(8), pp. 539-549.
- Maher, M. & Hughes, P., 1998. Recent developments in stochastic assignment modelling. *Traffic Engineering and Control*, 39(3), pp. 174-179..
- Mahmassani, H., Hu, T. & Jayakrishnan, R., 1992. *Dynamic Traffic Assignment and Simulation for Advanced Network Informatics*. Capri, Italy., Proceedings of the 2nd Annual Capri Seminar on Urban Traffic Networks.
- Mahmassani, H., Joseph, T. & Jou, R., 1993. Survey approach for study of urban commuter choice dynamics. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1412, pp. 80-89.

- Mahmassani, H. S., Caplice, C. G. & Walton, C. M., 1990. Characteristics of Urban Commuter Behavior: Switching Propensity and Use of Information. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1285, pp. 57-69.
- Mahut, M., Florian, M. & Tremblay, N., 2003. *Space-Time Queues and Dynamic Traffic Assignment: A model, algorithm and applications*. Washington, D.C., 82nd annual meeting of the Transportation Research Board.
- Mannering, F. & Kim, S.-G., 1994. Statistical analysis of commuters' route, mode, and departure time flexibility. *Transportation Research Part C*, 2(1), pp. 35-47.
- Mannering, F. L., 1989. Poisson analysis of commuter flexibility in changing routes and departure times. *Transportation Research Part B*, 23(1), p. 53-60.
- Manski, C., 1977. The structure of random utility models. *Theory and Decision*, Issue 8, pp. 229-254.
- Marchal, F., Hackney, J. & Axhausen, K. W., 2005. Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1935, pp. 93-100.
- McFadden, D., 1978. Modelling the choice of residential location. In: *Spatial interaction theory and residential location*. Amsterdam, North Holland: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J., pp. 75-96.
- McFadden, D., 2001. Economic choices. *The American Economic Review*, 91(3), pp. 351-378.
- McFadden, D. & Train, K., 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, Volume 15, pp. 447-470.
- Meloni, I., Spissu, E. & Bhat, C. R., 2011. The effect of personal cap and trade mileage policies on individual activity travel patterns: the Activity locator project. *Transportation Letters: The International Journal of Transportation Research*, 3(4), pp. 293-307.
- Merchant, D. & Nemhauser, G., 1978. A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science*, Volume 12, p. 183-199.
- Murakami, E. & Wagner, D., 1999. Can using Global Positioning System (GPS) improve trip reporting?. *Transportation Research Part C: Emerging Technologies*, 7(2-3), pp. 149-165.
- Nakazato, M. et al., 2006. Internet GIS-Based Activity-Travel Simulator for Investigating Alternative Activity-Travel Patterns. *11th International Conference on Travel Behaviour Research Kyoto*.
- Nielsen, O., 2000. A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B*, 34(5), pp. 377-402.

- Nielsen, O., 2004. Behavioural responses to pricing schemes: description of the Danish AKTA experiment. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 4(8), pp. 233-251.
- Nielsen, O., Daly, A. & Frederiksen, R., 2002. A stochastic multi-class road assignment model with distributed time and cost coefficients. *Networks and Spatial Economics*, Volume 2, pp. 327-346.
- Nielsen, O. & Frederiksen, R., 2006. Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Annals of Operations Research*, 144(1), pp. 263-285.
- Olshavsky, R., 1979. Task complexity and contingent processing in decision making: a replication and extension. *Organizational Behavior and Human Performance*, 24(3), pp. 300-316.
- Ortuzar, J. & Willumsen, L. G., 2001. *Modeling Transport*. s.l.:John wiley & sons.
- Papageorgiou, M., Blosseville, J. M. & H., H.-S., 1990. Modelling and real-time control of traffic flow on the southern part of boulevard périphérique in Paris: Part I: Modelling. *Transportation Research A*, Volume 24, pp. 345-359.
- Papinski, D. & Scott, D. M., 2010. A GIS-based toolokit for route choice analysis. *Journal of Transportation Geography*.
- Papinski, D., Scott, D. M. & Doherty, S. T., 2009. Exploring the route choice decision-making process: a comparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F*, 12(4), pp. 347-358.
- Parkany, E., Du, J., Aultman-Hall, L. & Gallagher, R., 2006. Modeling Stated and Revealed Route Choice: Consideration of Consistency, Diversion, and Attitudinal Variables. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1985, pp. 29-39.
- Park, D. & Rilett, L., 1997. Identifying multiple and reasonable paths in transportation networks: a heuristic approach.. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1607, pp. 31-37.
- Payne, H., 1971. Models of freeway traffic and control.. *Simulation Councils Proceeding Series: Mathematical Models of Public Systems*, Volume 1.
- Payne, J. W., 1976. Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behavior and Human Performance*, 16(2), pp. 366-387.

- Payne, J. W., Bettman, J. R. & Johnson, E. J., 1993. *The adaptive decision maker*. Cambridge, MA: Cambridge University Press.
- Peeta, S. & Ziliaskopoulos, A. K., 2001. Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future. *Networks and Spatial Economics*, Volume 1.
- Polydoropoulou, A., Ben-Akiva, M., Khattak, A. & Lauprete, G., 1996. Modeling Revealed and Stated En-Route Travel Response to Advanced Traveler Information Systems. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1537, pp. 38-45.
- Prashker, J. & Bekhor, S., 1998. Investigation of stochastic network loading procedures. *Transportation Research Record: Journal of the Transportation research Board*, Volume 1645, pp. 94-102.
- Prashker, J. & Bekhor, S., 2000. Congestion, stochastic, and similarity effects in stochastic user equilibrium models. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1733, pp. 80-87.
- Prashker, J. N. & Bekhor, S., 2004. Route choice models used in stochastic user equilibrium problem: a review. *Transport Reviews*, 24(4), pp. 437-463.
- Prato, C. & Bekhor, S., 2007. Modeling route choice behavior: how relevant is the choice set composition?. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 2003, pp. 64-73.
- Prato, C. G., 2009. Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2(1), pp. 65-100.
- Prato, C. G., 2012. Meta-analysis of choice set generation effects on route choice model estimates and predictions. *Transport*, 27(3), pp. 286-298.
- Prato, C. G. & Bekhor, S., 2006. Applying branch and bound technique to route choice set generation. *Transportation research record: Journal of the Transportation Research Board*, Volume 1985, pp. 19-28.
- Prato, C. G., Bekhor, S. & Pronello, C., 2005. Methodology for Exploratory Analysis of Latent Factors Influencing Drivers' Behavior. *Transportation Research Record*, Issue 1926, pp. 115-125.
- Prato, C. G., Bekhor, S. & Pronello, C., 2012. Latent variables and route choice behavior. *Transportation*, 39(2), pp. 299-319.

- Quddus, M. A., Ochieng, W. Y. & Noland, R. B., 2007. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, 15(5), pp. 312-328.
- Raghubir, P. & Krishna, A., 1996. As the crow flies: Bias in consumers' map-based distance judgments. *Journal of Consumer Research*, 23(1), pp. 26-39.
- Rakha, H. & Tawfik, A., 2009. Traffic Networks: Dynamic Traffic Routing, Assignment, and Assessment. In: *Encyclopedia of Complexity and Systems Science*. s.l.:s.n., pp. 9429-9470.
- Ramming, M. S., 2002. *Network Knowledge and Route Choice*. Cambridge: Ph.D. Thesis, Massachusetts Institute of Technology.
- Ramming, M. S., 2002. *Network Knowledge and Route Choice*. Cambridge: Ph.D. Thesis, Massachusetts Institute of Technology.
- Richardson, A., 1982. Search models and choice set generation. *Transportation Research Part A*, 16(5-6), pp. 403-419.
- Richards, P. I., 1956. Shockwaves on the Highway.. *Operations Research*, Volume 4, pp. 42-51.
- Rieser-Schüssler, N., Balmer, M. & Axhausen, K. W., 2013. Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science*, 9(9), pp. 825-845..
- Ruphail, N. et al., 1995. *A decision support system for dynamic pre-trip route planning*. Proceedings of The Fourth International Conference, Applications of Advanced Technologies in Transportation Engineering, pp. 325-329.
- Schönfelder, S. & Axhausen, K. W., 2004. Structure and innovation of human activity spaces. *Arbeitsbericht Verkehrs- und Raumplanung 258*, Zurich: IVT, ETH.
- Schuessler, N. & Axhausen, K. W., 2009. Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT). *Arbeitsberichte Verkehrs- und Raumplanung*, 568, IVT, ETH Zürich, Issue 568.
- Scott, K., Pabon-Jimenez, G. & Bernstein, D., 1997. *Finding alternatives to the best path*. Washington, D.C., Proceedings of the 76th Annual Meeting of the Transportation Research Board.
- Sheffi, Y., 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. s.l.:Prentice-Hall, (Englewood Cliffs, N.J.).
- Sheffi, Y. & Powell, W., 1982. An algorithm for the equilibrium assignment problem with random link times. *Networks*, Volume 12, pp. 191-207.

- Shier, D., 1976. Iterative methods for determining the K shortest paths in a network. *Networks*, Volume 6, pp. 205-229.
- Shier, D., 1979. On algorithms for finding the K-shortest paths in a network. *Networks*, Volume 9, pp. 195-214.
- Simon, H., 1955. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1), pp. 99-118.
- Spissu, E., Meloni, I. & Sanjust, B., 2011. Behavioral Analysis of Choice of Daily Route with Data from Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2230, pp. 96-103.
- Srinivasan, K. & Mahmassani, H., 2003. Analyzing heterogeneity and unobserved structural effects in route-switching behavior under ATIS: a dynamic kernel logit formulation. *Transportation Research Part B: Methodological*, 37(9), pp. 793-814.
- Stern, E., Holm, E. & Maarseven, M., 1993. Information and Commuters Behaviour: A comparative analysis. *Europe on the Move*, pp. 350-369.
- Stopher, P., 1980. Captivity and choice in travel-behavior models. *Transportation Journal of ASCE*, Volume 106, pp. 427-435.
- Stopher, P., Kockelman, K., Greaves, S. & Clifford, E., 2008. Reducing burden and sample sizes in multiday household travel surveys. *Transportation Research Record: Journal of the Transportation Research Board*, 2064(1), pp. 12-18.
- Stopher, P. R. & Greaves, S. P., 2007. Household travel surveys: Where are we going?. *Transportation Research Part A* 41, p. 367-381.
- Swait, J. & Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B - Methodological*, 21(2), pp. 91-102.
- Tawfik, A. M., Rakha, H. A. & Miller, S. D., 2010. An Experimental Exploration of Route Choice: Identifying Drivers Choices and Choice Patterns and Capturing Network Evolution. *13th International IEEE, Annual Conference on Intelligent Transportation Systems*.
- Tawfik, A., Rakha, H. & Miller, S., 2010. *Driver route choice behavior: Experiences, perceptions, and choices*. s.l., Intelligent Vehicles Symposium (IV), 2010 IEEE, pp. 1195 - 1200 .
- Train, K., 2009. *Discrete Choice Methods with Simulation*. 2nd ed. 2009: Cambridge University Press.
- Train, K. & Sonnier, G., 2005. Mixed logit with bounded distributions of correlated partworths. *Applications of Simulation Methods in Environmental and Resource Economics*, pp. 117-134.



- Tversky, A., 1969. Intransitivity of preferences. *Psychological Review*, Volume 76, pp. 31-48.
- Tversky, A., 1972. Choice by elimination. *Journal of Mathematical Psychology*, Volume 9, pp. 341-367.
- Tversky, A., 1972. Elimination by aspects: a theory of choice. *Psychological Review*, Volume 79, pp. 281-299.
- Tversky, A. & Kahneman, D., 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, Volume 9, pp. 195-230.
- Van Aerde, M., 1994. *INTEGRATION: A model for simulating integrated traffic networks*, Kingston, Canada.: Transportation Systems Research Group, Queen's University.
- Van der Zij, N. & Fiorenzo-Catalano, S., 2005. Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B*, 39(6), pp. 545-563.
- Vovsha, P., 1997. The cross-nested logit model: application to mode choice in the Tel Aviv metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1607, pp. 13-20.
- Wardrop, J., 1952. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers Part 2 1*, pp. 325-378.
- Wen, C. & Koelman, F., 2001. The generalized nested logit model. *Transportation Research Part B*, 35(7), pp. 627-641.
- White, C. E., Bernstein, D. & Kornhauser, A. L., 2000. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8(1), pp. 91-108.
- Wickens, C. & Hollands, J., 1999. *Engineering Psychology and Human Performance (3rd Edition)*. s.l.:Prentice Hall.
- Williams, H., 1977. On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning*, Volume 9A, pp. 285-344.
- Williams, H. & Ortuzar, J., 1982. Behavioural theories of dispersion and the misspecification of travel demand models. *Transportation Research Part B: Methodological*, 16(3), p. 167-219.
- Wolf, J. et al., 2014. *Applying GPS Data to Understand Travel Behavior*, Washington, D.C.: Transportation Research Board of the National Academies.

- Wolf, J., Bricka, S., Ashby, T. & Gorugantua, C., 2004. *Advances in the Application of GPS to Household Travel Surveys*. Washington, D.C., Presented at the National Household Travel Survey Conference.
- Wolf, J. et al., 2004. Eighty weeks of global positioning system traces: approaches to enriching trip information. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 1870, pp. 46-54.
- Xu, C., Wang, W., Li, Z. & Yang, C., 2010. *Comparative study on drivers' route choice response to travel information at different departure time*. s.l., CAR'10 Proceedings of the 2nd international Asia conference on Informatics in control, automation and robotics, pp. 97-100.
- Xu, H., Zhou, J. & Xu, W., 2011. A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. *Transportation Research Part C*, Volume 19, pp. 218-228.
- Yai, T., Iwakura, S. & Morichi, S., 1997. Multinomial probit with structured covariance for route choice behavior. *Transportation Research Part B*, 31(3), pp. 195-207..
- Yen, J., 1971. Finding the K shortest loopless paths in a network. *Management Science*, Volume 17, pp. 712-716.
- Zhang, L. & Levinson, D., 2008. Determinants of Route Choice and Value of Traveler Information: A Field Experiment. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2086, pp. 81-92.
- Zhou, J. (. & Golledge, R., 2007. Real-time tracking of activity scheduling/schedule execution within a unified data collection framework. *Transportation Research Part A* 41, p. 444-463.
- 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*.
- Zhou, Z. & Chen, A., 2003. *Stochastic user equilibrium problem: a comparison between length-based and congestion-based C-Logit models*. Hong Kong, China, Proceedings of the 8th Hong Kong Society of Transportation Studies Conference: Transportation and Logistics, pp. 244-253.
- Zhu, S. & Levinson, D., 2010. *Do people use the shortest path? An empirical test of Wardrop's first principle*. Minneapolis, s.n.

