



Università degli studi ROMA TRE

Scuola dottorale in Ingegneria
Dottorato di ricerca in Scienze dell'Ingegneria Civile
XXI Ciclo

Tesi di Dottorato

Correzione della domanda di trasporto in
dinamica intraperiodale con l'ausilio di differenti
fonti di dati

Dottorando: Marialisa Nigro

Docente guida: Prof. Stefano Gori

Coordinatore del dottorato: Prof. Leopoldo Franco

Roma, Febbraio 2009

Collana delle tesi di Dottorato di Ricerca
In Scienze dell'Ingegneria Civile
Università degli Studi Roma Tre
Tesi n° 23

Sommario

La domanda di trasporto è un dato essenziale nei modelli di simulazione del traffico; proprio dalla sua interazione con l'offerta è possibile definire, tramite il processo di assegnazione, il flusso che interessa i diversi elementi che compongono il sistema di trasporto in analisi e procedere quindi ad una valutazione delle condizioni di deflusso che si verificano sulla rete.

Effettuare una correzione della domanda di trasporto significa determinare l'entità e la distribuzione degli spostamenti tra le diverse zone dell'area di studio, in un determinato intervallo temporale, per differenti scopi, con le diverse modalità di trasporto disponibili, partendo da una stima iniziale della variabile in esame (la domanda di trasporto) ed aggiornando tale stima attraverso l'uso di una o più fonti di dati.

Qualora le fonti di dati siano più di una, o in generale di differente natura, si può parlare anche di stima combinata della domanda di trasporto.

In particolare i dati in gioco sono usualmente pregresse informazioni sulla domanda di trasporto (dati censuari e dati da modello), nonché misure di tipo aggregato raccolte direttamente sulla rete (flussi di traffico).

Quando la domanda, così come l'offerta, si mantiene costante per un periodo di tempo sufficientemente lungo da permettere al sistema analizzato di raggiungere una condizione di stazionarietà, si è di fronte a modelli cosiddetti di tipo statico; nel caso in cui invece la domanda vari all'interno del periodo temporale di riferimento, con variazioni che conducono a fenomeni di punta di durata non sufficiente a raggiungere una distribuzione stabile dei flussi sulla rete, si rendono necessari i cosiddetti modelli dinamici in cui viene assunta l'ipotesi di dinamica intraperiodale.

Nel passaggio dal contesto statico al contesto dinamico, il problema di correzione della domanda di trasporto aumenta di difficoltà sia a causa dell'aumento della dimensione stessa del problema, che per la necessità imprescindibile di riprodurre le effettive condizioni di congestione verificatesi sulla rete.

Il presente lavoro propone a tal fine un nuovo metodo per effettuare la correzione della domanda di trasporto in dinamica intraperiodale che possa portare ad un risultato affidabile in tempi computazionali accettabili.

Il metodo è basato su un approccio in simulazione ed utilizza come informazioni di partenza misure aggregate quali flussi e velocità veicolari (in grado così di determinare univocamente lo stato – congestionato o non congestionato – dei singoli elementi della rete), nonché precedenti informazioni sulla domanda sia in termini di singoli spostamenti origine/destinazione che di tipo più aggregato (spostamenti generati da ogni singola zona). Data la differente natura dei dati utilizzati, nel seguito si userà spesso la definizione di “stima combinata” come sinonimo del problema di “correzione della domanda”.

La procedura risolutiva proposta è una variante del Simultaneous Perturbation Stochastic Approximation (SPSA, Spall 1998-1999), metodo basato su un'approssimazione del gradiente, attraverso la perturbazione simultanea di ogni variabile, ed in grado di determinare una buona soluzione qualora il punto di partenza sia sufficientemente vicino alla soluzione ottima (ricerca locale). Le modifiche apportate all'algoritmo SPSA riguardano sia la fase di calcolo della stima del gradiente che la ricerca del minimo lungo la direzione di discesa e permettono di ridurre ulteriormente gli oneri computazionali della procedura e di evitare la calibrazione da caso a caso dei coefficienti presenti nelle legge proposta originariamente per il calcolo del passo.

La possibilità di applicare la procedura nel caso di reti di dimensioni reali è senz'altro l'aspetto più importante del presente lavoro di ricerca; di fatti gli esempi applicativi presenti in letteratura fanno sempre riferimento a casi molto semplici e qualora si cerchi il passaggio verso contesti reali, ci si scontra con tempi computazionali eccessivamente elevati.

La fase dello studio che vede l'applicazione al contesto reale della città di Calgary è stata svolta presso il “Centre de recherche sur les transports” dell'Università di Montreal con la supervisione del prof. Michael Florian e dell'Ing. Michael Mahut della INRO Consultants.

Abstract

Transport demand is an essential input to traffic simulation models; the interaction between transport demand and supply, known as the assignment model, permits to identify flows for each element on the network and so to evaluate the traffic conditions.

Perform a demand adjustment consists in determining the entity and the distribution of trips between each couple of zones in the study area, for a prefixed time interval, for different mobility reasons and with different transport modes, starting from a previous demand estimate value and uploading it using different types of information.

If more than one type of information is used, it is possible to define this problem also as a combined demand estimation problem.

The information used are usually prior demand data (census data or model data) and traffic measures collected on the network (traffic flows).

If the demand varies during the observation interval with peak variations that don't permit to reach a steady-state of flows on the network, we have a dynamic model with dynamic intra-periods hypothesis.

Passing from the static to the dynamic context, the demand adjustment problem increases its complexity due to the larger dimension of the problem, that to the need to reflect the congestion conditions on the network.

This work presents a new method to solve the adjustment of dynamic traffic demand, searching for a reliable solution with acceptable computational times. It uses as an input traffic counts and speeds (to identify the state of each link – congested or uncongested), a-priori origin/destination matrices and other aggregate demand data (traffic demand productions by zone). For the different types of information used, in the following “combined estimation” will be used as a synonymous of “demand adjustment”.

The proposed solving procedure is a modification of the basic Simultaneous Perturbation Stochastic Approximation (SPSA, Spall 1998-1999) path search optimization method; it can find a good solution when the starting point (the seed matrix) is assumed to be “near” the optimal one (local search), working with a gradient approximation based on a simultaneous perturbation of each demand variable. Modifications to the

SPSA algorithm concern both the gradient approximation and the computation of new solution along the descendent direction: they are proposed in order to reduce computational times and to avoid calibration of parameters inside the basic SPSA step rule formulation (for the line search along the descent direction).

The chance to apply the resulting procedure on real networks is the most important goal of this research; in literature other applications to the same problem refer to simple cases and if a real context is analyzed, computational times increase a lot.

The study phase regarding the application on the real network of Calgary has been conducted at the “Centre de recherche sur les transports” of Montreal University with the supervision of prof. Michael Florian and Dr. Michael Mahut (INRO Consultants).

Indice

ELENCO DELLE FIGURE.....	IX
ELENCO DELLE TABELLE	XII
ELENCO DEI SIMBOLI	XIV
1. INTRODUZIONE	1
2. LA STIMA DELLA DOMANDA DI TRASPORTO	5
2.1 LA DOMANDA DI TRASPORTO PER LA SIMULAZIONE DEL TRAFFICO	5
2.1.1 LA SIMULAZIONE STATICA E DINAMICA DEL TRAFFICO	6
2.2 LA STIMA DELLA DOMANDA DI TRASPORTO.....	10
2.2.1 LA STIMA DIRETTA DELLA DOMANDA DI TRASPORTO ATTUALE	11
2.2.2 LA STIMA DISAGGREGATA DEI MODELLI DI DOMANDA	12
2.3 PROBLEMA DI CORREZIONE DELLA MATRICE O/D TRAMITE CONTEGGI DI TRAFFICO (CASO STATICO)	13
2.3.1 STIMATORI O MISURE DI GOODNESS OF FIT	16
2.4 PROBLEMA DI CORREZIONE DELLA MATRICE O/D TRAMITE CONTEGGI DI TRAFFICO (CASO DINAMICO)	20
2.4.1 APPROCCIO DI ASHOK	21
2.4.2 APPROCCIO DI VAN DER ZIJP	22
2.4.3 APPROCCIO DI BALAKRISHNA.....	23
2.5 OSSERVAZIONI.....	24
3. FORMALIZZAZIONE DEL PROBLEMA.....	27
3.1 OBIETTIVI.....	27
3.2 FUNZIONE OBIETTIVO E VINCOLI.....	28
3.3 LE CARATTERISTICHE DEL PROBLEMA DI CORREZIONE DELLA DOMANDA (CASO DINAMICO)	31
3.4 METODI DI OTTIMIZZAZIONE PER PROBLEMI A LARGA SCALA, NON LINEARI E STOCASTICI	32
3.4.1 METODI PATH SEARCH.....	33
3.4.2 METODI PATTERN SEARCH.....	36
3.4.3 METODI RANDOM SEARCH.....	37
3.5 PROCEDURA RISOLUTIVA.....	39
3.5.1 AGGIORNAMENTO DELLA SOLUZIONE CORRENTE (SPALL, GSM, PI).....	42
3.5.2 GRADIENTE APPROSSIMATO (SD – AD).....	46
3.6 STRUMENTI DI RICERCA (DYNAMEQ).....	46
3.6.1 RAPPRESENTAZIONE DELLA RETE, SISTEMI DI CONTROLLO E DOMANDA.....	47
3.6.2 IL MODELLO DI ASSEGNAZIONE DINAMICA.....	48
3.6.3 L'ALGORITMO RISOLUTIVO.....	49

3.6.4	IL MODELLO DI CARICAMENTO DELLA RETE.....	53
4.	APPLICAZIONE RETE TEST	57
4.1	ANALISI DI SENSITIVITÀ.....	66
5.	APPLICAZIONE RETE DI CALGARY	73
5.1	ESPERIMENTI CON UN INTERVALLO TEMPORALE	73
5.2	ESPERIMENTI CON QUATTRO INTERVALLI TEMPORALI	79
5.2.1	INFLUENZA DEL TIPO E DELLA NUMEROSITÀ DELLE INFORMAZIONI SULLA PROCEDURA DI CORREZIONE	82
5.2.2	INTERVALLO DI SIMULAZIONE E DIMENSIONE DELLA DOMANDA	89
5.2.3	INTRODUZIONE DI VINCOLI PIÙ RESTRITTIVI SULLA DOMANDA.....	94
5.2.4	DIMENSIONE DEL PROBLEMA E TEMPI COMPUTAZIONALI.....	97
6.	CONCLUSIONI E FUTURI SVILUPPI.....	101
7.	BIBLIOGRAFIA	105

Elenco delle figure

2.1	<i>Rappresentazione matriciale della domanda di trasporto (caso statico e caso dinamico)</i>	7
2.2	<i>Simulazione statica e dinamica</i>	8
2.3	<i>Modello di assegnazione (a) e problema di stima della domanda di trasporto (b)</i>	13
2.4	<i>Indeterminatezza del problema di stima della domanda da conteggi di traffico</i>	15
2.5	<i>MSA per la risoluzione del problema di stima O/D</i>	16
2.6	<i>Diagramma fondamentale parabolico e triangolare ed individuazione dello stato di congestione</i>	26
3.1	<i>Symmetric e Asymmetric Design</i>	34
3.2	<i>Confronto tra la convergenza di SPSA e FDSA</i>	36
3.3	<i>Algoritmo SPSA applicato alla correzione della matrice OD</i>	41
3.4	<i>Rete test</i>	42
3.5	<i>Andamento della funzione obiettivo per rete test in condizioni di domanda e offerta semplificate utilizzando lo step classico</i>	43
3.6	<i>Applicazione del metodo della sezione aurea per rete test in condizioni di domanda e offerta semplificate</i>	43
3.7	<i>Confronto funzione obiettivo con utilizzo step classico e GSM</i>	44
3.8	<i>Il metodo della sezione aurea (GSM)</i>	45
3.9	<i>Applicazione dell'interpolazione polinomiale per il calcolo</i>	45

della soluzione lungo la direzione di discesa

3.10	<i>Struttura dell'algoritmo risolutivo del modello di assegnazione dinamica</i>	50
3.11	<i>Algoritmo di assegnazione dinamica all'equilibrio in Dynameq</i>	52
3.12	<i>Diagramma triangolare flusso-densità</i>	53
3.13	<i>Modello semplificato di veicolo accodato</i>	54
4.1	<i>Andamento della funzione obiettivo (2 intervalli temporali)</i>	59
4.2	<i>Corrispondenze conteggi - flussi simulati (2 intervalli temporali)</i>	60
4.3	<i>Corrispondenze velocità misurate-simulate (2 intervalli temporali)</i>	61
4.4	<i>Corrispondenze domanda stimata-reale (2 intervalli temporali)</i>	61
4.5	<i>Andamento della funzione obiettivo (5 intervalli temporali)</i>	62
4.6	<i>Corrispondenze conteggi - flussi simulati (5 intervalli temporali)</i>	64
4.7	<i>Corrispondenze velocità misurate-simulate (5 intervalli temporali)</i>	65
4.8	<i>Corrispondenze domanda stimata-reale (5 intervalli temporali)</i>	65
4.9	<i>Corrispondenza in termini di flussi e velocità senza vincolo di generazione</i>	71
5.1	<i>Rete di Calgary in Dynameq</i>	73
5.2	<i>Andamento temporale della domanda "reale"</i>	79

5.3	<i>Confronto tra domanda stimata e domanda reale</i>	82
5.4	<i>Scattergramm flussi rilevati/flussi misurati (Tabella 5.10 ipotesi di domanda alta)</i>	84
5.5	<i>Scattergramm velocità stimate/velocità misurate (Tabella 5.10 ipotesi di domanda alta)</i>	84
5.6	<i>Flussi e velocità di manovra alle intersezioni</i>	85
5.7	<i>Flussi e velocità delle manovre di svolta a sinistra alle intersezioni</i>	86
5.8	<i>Interazione tra domanda e vincoli di capacità della rete</i>	90
5.9	<i>Scattergramm flussi/velocità rilevati/e vs flussi/velocità misurati/e (Calgary scaricamento)</i>	93
5.10	<i>Scattergramm matrice “reale” vs matrice stimata (Calgary scaricamento)</i>	94
5.11	<i>Scattergramm matrice “reale” vs matrice stimata (Calgary scaricamento con uguaglianza generazioni)</i>	96
5.12	<i>Scattergramm matrice “reale” vs matrice stimata (Calgary senza scaricamento con uguaglianza generazioni)</i>	96
5.13	<i>Applicazione alla Sezione autostradale I-880 California (Ashok, 1996)</i>	98
5.14	<i>Applicazione “Waterfront corridor”, Toronto (Kattan e Abdulhai, 2005)</i>	99
5.15	<i>Applicazione alla rete di Los Angeles con 33 zone “attive” (Balakrishna, 2006)</i>	99

Elenco delle tabelle

2.1	<i>Definizione degli stimatori per le differenti teorie adottate ed ipotesi sui residui</i>	19
4.1	<i>Differenti ipotesi testate del SPSA sulla rete test</i>	57
4.2	<i>Confronto tra le procedure (2 intervalli temporali)</i>	59
4.3	<i>Valori iniziali degli indicatori R^2 ed RME</i>	59
4.4	<i>Risultati a convergenza (2 intervalli temporali)</i>	60
4.5	<i>Confronto tra le procedure (5 intervalli temporali)</i>	63
4.6	<i>Valori iniziali degli indicatori R^2 ed RME</i>	63
4.7	<i>Risultati a convergenza (5 intervalli temporali)</i>	63
4.8	<i>Risultati dell'analisi di sensitività (R^2 ed MSE)</i>	67
4.9	<i>Risultati dell'analisi di sensitività (RME ed RRMSE)</i>	68
4.10	<i>Funzione Obiettivo con pesi (R^2 e MSE)</i>	70
4.11	<i>Funzione Obiettivo con pesi (RME e RRMSE)</i>	70
4.12	<i>Introduzione del vincolo di generazione (R^2 e MSE)</i>	71
4.13	<i>Introduzione del vincolo di generazione (RME e RRMSE)</i>	72
5.1	<i>Valori iniziali della funzione obiettivo e dell'RME - Range di perturbazione: 15÷25%</i>	75
5.2	<i>Valori iniziali della funzione obiettivo e dell'RME - Range di perturbazione: 25÷35%</i>	76
5.3	<i>Risultati con vincolo di generazione - Range di perturbazione: 15÷25%</i>	76

5.4	<i>Risultati con vincolo di generazione - Range di perturbazione: 25÷35%</i>	76
5.5	<i>Risultati senza vincolo di generazione - Range di perturbazione: 15÷25%</i>	77
5.6	<i>Risultati senza vincolo di generazione - Range di perturbazione: 25÷35%</i>	77
5.7	<i>Risultati - Normalizzazione dei termini della funzione obiettivo</i>	78
5.8	<i>Valori iniziali della funzione obiettivo e dell'RME (Calgary 4 intervalli temporali)</i>	80
5.9	<i>Risultati con vincolo di generazione (Calgary 4 intervalli temporali)</i>	80
5.10	<i>Risultati con vincolo di generazione e normalizzazione (Calgary 4 intervalli temporali)</i>	81
5.11	<i>Risultati senza velocità di manovra (Calgary 4 intervalli temporali)</i>	87
5.12	<i>Risultati senza velocità di manovra con incremento sezioni d'arco (Calgary 4 intervalli temporali)</i>	88
5.13	<i>Stima dell'errore per differenti valori delle variabili</i>	92
5.14	<i>Risultati con scaricamento della rete (Calgary)</i>	94
5.15	<i>Risultati con uguaglianza delle generazioni (Calgary)</i>	97

Elenco dei simboli

Nell'elenco che segue sono riportati i principali simboli che compaiono nei capitoli della tesi.

A	<i>Archi direzionali della rete di trasporto</i>
N	<i>Nodi della rete di trasporto</i>
S	<i>Sottoinsieme di archi forniti di sensori</i>
T	<i>Periodo di analisi</i>
n_h	<i>Numero di intervalli temporali in cui è suddiviso il periodo di analisi</i>
n_{od}	<i>Numero di coppie origine-destinazione</i>
n_s	<i>Numero di sensori sulla rete</i>
\mathbf{x}_i	<i>Matrice di domanda stimata per l'intervallo di partenza i</i>
\mathbf{d}_i	<i>Matrice di domanda target per l'intervallo di partenza i</i>
\mathbf{v}_i	<i>Flussi simulati sugli archi appartenenti ad S nell'intervallo i</i>
$\hat{\mathbf{v}}_i$	<i>Conteggi di traffico sugli archi appartenenti ad S nell'intervallo i</i>
\mathbf{s}_i	<i>Velocità simulate sugli archi appartenenti ad S nell'intervallo i</i>
$\hat{\mathbf{s}}_i$	<i>Velocità misurate sugli archi appartenenti ad S nell'intervallo i</i>
G_o^*	<i>generazione da modello della zona o</i>
G_o^i	<i>Generazione da ricostruzione della zona o nell'intervallo i</i>
$\hat{\mathbf{g}}$	<i>Stima del gradiente</i>

Δ *vettore di perturbazioni random*

$\hat{\mathbf{g}}$ *Gradiente approssimato medio*

1. Introduzione

La domanda di trasporto viene definita come l'insieme degli spostamenti che avvengono tra le diverse zone di una determinata area di studio, per differenti motivi, in diversi periodi temporali, con le diverse modalità di trasporto disponibili (Cascetta, 1998).

La domanda di trasporto risulta di fondamentale importanza nei modelli di simulazione del traffico. Difatti sia l'analisi che la progettazione di interventi sui sistemi di trasporto richiedono la conoscenza della domanda attuale e/o la previsione di quella futura: solo dalla interazione domanda-offerta è possibile definire, tramite il processo di assegnazione, le caratteristiche del deflusso (velocità, densità e flusso) che interessano i diversi elementi infrastrutturali presenti nel sistema di analisi, quindi valutare il livello di servizio ed i costi totali del trasporto necessari per la fase progettuale.

Ovviamente, essendo la domanda futura una proiezione attraverso modelli matematici della domanda attuale, maggiore sarà la bontà di quest'ultima e maggiore sarà la bontà della valutazione del sistema nello stato futuro. Risulta dunque necessario partire sempre da una buona stima di quella che è la matrice di domanda all'orizzonte temporale attuale.

Si parla di dinamica intraperiodale quando la domanda varia all'interno del periodo di riferimento, con variazioni che conducono a fenomeni di punta di durata non sufficiente a raggiungere una distribuzione stabile dei flussi sulla rete. In questo caso l'usuale rappresentazione a singola matrice del caso statico, lascia il posto ad una rappresentazione ad n matrici, tante quante gli intervalli temporali in cui è stato suddiviso il periodo di riferimento T .

L'approccio utilizzato per stimare la domanda nel caso dinamico è solitamente una estensione dell'approccio statico che vede la minimizzazione tra domanda stimata e domanda iniziale nota (detta matrice target), nonché della distanza tra flussi d'arco derivanti dall'assegnazione della matrice stimata e conteggi di traffico disponibili su di un sottoinsieme di archi della rete. Per la presenza di una stima iniziale (matrice target) della variabile in esame si può parlare, più che di

sola stima, di correzione della domanda di trasporto attraverso l'ausilio di differenti fonti di dati.

Se le informazioni relative ai conteggi di traffico possono risultare soddisfacenti nel caso statico, non lo sono più quando si considera la dinamica. Nei modelli dinamici vengono simulate esplicitamente la formazione ed il dissipamento delle code, i colli di bottiglia ed i fenomeni di spill-back e assumono in tale frangente un'importanza rilevante grandezze quali densità e velocità, le sole che possono discriminare tra il trovarsi in uno stato congestionato o non congestionato. Difatti il flusso da solo non è in grado di dare informazioni sullo stato del sistema.

Ulteriori problematiche legate al processo di stima nel caso dinamico derivano sempre dalla dimensione temporale del problema che permette di distinguere tra un approccio simultaneo (individuare le matrici per ogni intervallo temporale in cui è suddiviso il periodo di analisi T , che minimizzano una determinata misura della distanza tra domanda stimata e domanda iniziale nota (matrici target), nonché della distanza tra flussi d'arco derivanti dall'assegnazione della matrice stimata e conteggi di traffico) ed un approccio sequenziale (individuare una matrice alla volta, partendo dal primo intervallo temporale e procedendo sino all'ultimo, mantenendo di volta in volta fisse le matrici precedentemente calcolate).

In quest'ultimo caso si ottiene una riduzione della complessità computazionale del problema in quanto il problema stesso viene ad essere scomposto in sottoproblemi più semplici da trattare ed ogni matrice stimata può essere utilizzata come stima iniziale per l'intervallo temporale successivo.

Ovviamente l'approccio sequenziale è richiesto nel caso di stima della domanda in un contesto dinamico di tipo on-line, dove l'aggiornamento in tempo reale dei valori dei flussi OD avviene tramite le misure di traffico disponibili mano a mano che i sistemi di rilevamento acquisiscono il dato e dunque in un contesto in cui i tempi computazionali diventano un vincolo essenziale del processo di stima.

Nel caso si voglia ricorrere ad un approccio simultaneo, dunque per una stima off-line, formalmente più corretto e perciò in grado di restituire un'informazione più accurata sulla domanda è da tener presente la mancata possibilità di utilizzare le matrici di assegnazione per esprimere il legame flussi-domanda, in quanto è dimostrato che l'approccio simultaneo basato sulle matrici di assegnazione risulta avere una complessità computazionale proibitiva anche su reti di medie dimensioni

(Cascetta e Russo, 1997; Toledo et al., 2003; Bierlaire e Crittin, 2004). Ancor più, qualora si vogliano inserire ulteriori misure di traffico come velocità e densità all'interno del processo di stima, è da tener conto la non linearità delle relazioni esistenti tra flusso, domanda e le suddette variabili e la mancata esistenza del legame tramite matrici di assegnazione.

La presente tesi si pone l'obiettivo di risolvere il problema di correzione/stima combinata della domanda di trasporto attuale in dinamica intraperiodale per un contesto di tipo off-line.

Si ipotizza di partire da una serie di matrici target note, una per ogni intervallo temporale che costituisce il periodo di riferimento e di correggere tali matrici attraverso l'utilizzo di misure di traffico.

E' sempre possibile ipotizzare di partire da una precedente informazione sulla domanda sufficientemente buona ed il problema riveste in tal senso l'ambito di una ricerca locale, per cui anche la scelta del metodo risolutivo adottato deriva dalle ipotesi effettuate circa la bontà del punto di partenza (ovvero dalla bontà delle matrici target).

La soluzione a cui si vuole giungere è una soluzione affidabile, ovvero si vuole dimostrare che la domanda stimata non comporta solo una buona corrispondenza con le misure di traffico, ma che sia effettivamente rappresentativa degli spostamenti effettuati dagli utenti della rete nei diversi intervalli temporali. Per far ciò è necessario puntare su un approccio di tipo simultaneo che possa cogliere l'influenza di ogni matrice sui diversi intervalli temporali e considerare ulteriori elementi oltre i flussi che possano rappresentare al meglio le effettive condizioni di congestione della rete.

La tesi è così strutturata: a seguito della presente introduzione, nel Capitolo 2 viene definita la domanda di trasporto soffermandosi sulle diverse rappresentazioni assunte da questa nel caso statico e nel caso dinamico e sottolineandone l'importanza all'interno dei modelli matematici dell'Ingegneria dei Trasporti. Si passa quindi alla definizione del problema di stima della domanda con una rassegna degli approcci presenti in letteratura e delle problematiche che da essi scaturiscono.

Nel Capitolo 3 viene formalizzato il modello adottato per affrontare il problema di correzione/stima combinata della domanda in dinamica intraperiodale. Si passa quindi alla ricerca del metodo risolutivo proponendo varianti che possano accrescere l'efficienza degli algoritmi solitamente utilizzati per risolvere tali tipologie di problemi.

Nei Capitoli 4 e 5 vengono infine effettuate le applicazioni del modello rispettivamente su di una rete test e su una porzione della rete di Calgary in Canada.

2. La stima della domanda di trasporto

2.1 *La domanda di trasporto per la simulazione del traffico*

La domanda di trasporto viene definita come l'insieme degli spostamenti che avvengono tra le diverse zone di una determinata area di studio, per differenti motivi, in diversi periodi temporali, con le diverse modalità di trasporto disponibili (Cascetta, 1998).

Gli spostamenti, così come le caratteristiche che li competono, derivano dalla necessità di svolgere attività in luoghi diversi e sono così direttamente collegati al sistema delle attività nonché al sistema dell'offerta di trasporto di cui usufruiscono. Tale relazione può essere espressa in termini formali come:

$$d[K_1, K_2, \dots, K_n] = d(\mathbf{SE}, \mathbf{T}, \boldsymbol{\beta}) \quad (2.1)$$

dove

K_1, K_2, \dots, K_n = caratteristiche del flusso medio di spostamenti;

\mathbf{SE} = vettore delle variabili socioeconomiche relative al sistema delle attività;

\mathbf{T} = variabili di livello di servizio relativo al sistema dell'offerta;

$\boldsymbol{\beta}$ = vettore di coefficienti incogniti determinabili tramite calibrazione della funzione di domanda.

Quando si parla di sistema delle attività si intende l'unione di tre sottosistemi: le residenze, le superfici con la loro rispettiva destinazione d'uso (es. industriale, agricola, terziaria) e le attività economiche presenti. E' evidente come la distribuzione delle famiglie e delle attività in una data area di studio determini scelte di mobilità e di viaggio (quando spostarsi, dove spostarsi, con quale mezzo, etc.). Nella funzione in (2.1) compare anche il sistema dell'offerta; quest'ultimo si definisce come l'insieme delle infrastrutture e dei servizi necessari a soddisfare il bisogno di spostarsi, di fatti tutte le componenti dello spostamento vengono ad essere influenzate dalle caratteristiche dei sistemi di trasporto presenti tra le

diverse zone dell'area di studio. E' sufficiente pensare a come la scelta della destinazione dipenda anche dal costo generalizzato di trasporto per raggiungerla o come il modo di trasporto venga stabilito anche in virtù di tempi, costi e affidabilità di modi alternativi.

La determinazione della domanda di trasporto appare immediatamente come un'operazione di non semplice sviluppo per la presenza di una serie di elementi che concorrono nel determinarla e per la tipologia di legame esistente tra questi elementi, per lo più non lineare e con la presenza di numerosi cicli di retro-azione.

La rappresentazione classica della domanda di trasporto è di tipo matriciale e viene realizzata attraverso la suddivisione degli spostamenti relativi ad un determinato intervallo temporale tra le diverse origini e destinazioni (matrice O-D) dell'area di studio. Tale matrice ha un numero di righe e di colonne pari al numero di zone tra le quali possono avvenire gli spostamenti ed il generico elemento d_{od} contenuto nella cella della matrice fornisce il numero di spostamenti che hanno origine dalla zona o e destinazione nella zona d nell'unità di tempo considerata (Figura 2.1 - a).

La domanda di trasporto risulta di fondamentale importanza nei modelli di simulazione del traffico; infatti proprio dalla sua interazione con l'offerta è possibile definire, tramite il processo di assegnazione, il flusso che interessa i diversi elementi infrastrutturali presenti nel sistema di analisi e quindi il dato principale per una seguente valutazione progettuale.

2.1.1 La simulazione statica e dinamica del traffico

La variabile "tempo" è una delle caratteristiche principali nella definizione della domanda: i modelli matematici che descrivono i sistemi di trasporto considerano usualmente la domanda, così come l'offerta, costante per un periodo di tempo sufficientemente lungo da permettere al sistema analizzato di raggiungere una condizione di stazionarietà: si è di fronte in questo caso a modelli cosiddetti di tipo statico (Figura 2.1 - a).

Qualora si consideri la domanda variabile all'interno del periodo di riferimento, con variazioni che conducono a fenomeni di punta di durata non sufficiente a raggiungere una distribuzione stabile dei flussi sulla rete,

si passa a modelli dinamici in cui viene assunta l'ipotesi di dinamica intraperiodale.

In quest'ultimo caso la rappresentazione matriciale vedrà la presenza di un numero n di matrici (Figura 2.1 - b), tante quanti gli intervalli temporali in cui è stato suddiviso il periodo di riferimento T : il generico elemento d'_{od} contenuto nella cella della matrice relativa all'intervallo t fornisce così il numero di spostamenti in partenza dalla zona o verso la zona d nell'intervallo t .

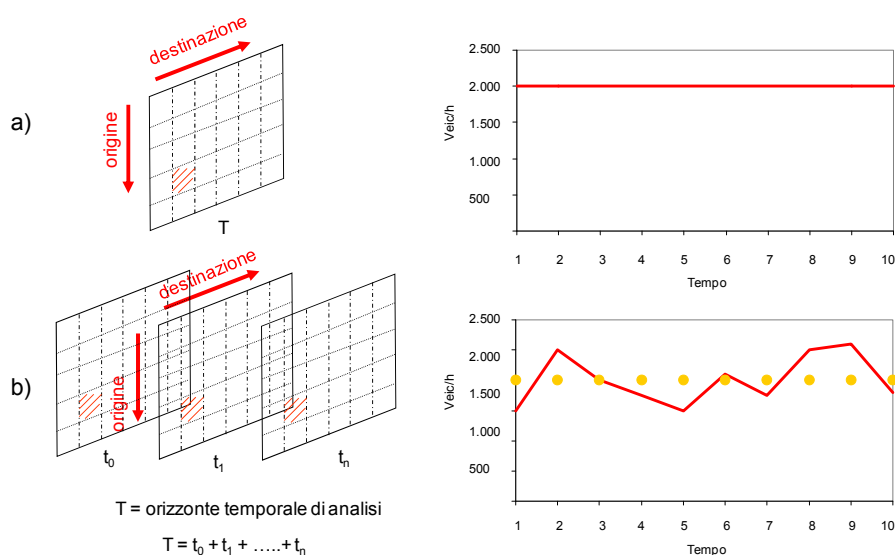


Figura 2.1: Rappresentazione matriciale della domanda di trasporto (caso statico e caso dinamico)

Nel caso statico, il processo di assegnazione (Figura 2.2 - a) fornisce dei valori di flusso costanti nell'orizzonte temporale di analisi; viceversa, nel caso dinamico i flussi risultanti sono variabili nel tempo (Figura 2.2 - b). Non è questa però l'unica differenza tra una simulazione statica ed una simulazione dinamica del traffico. Difatti la simulazione dinamica nasce per superare una serie di limiti del processo statico, inadeguato nel rappresentare esplicitamente i vincoli di capacità sulla rete e di fornire informazioni sulla propagazione della congestione nel tempo.

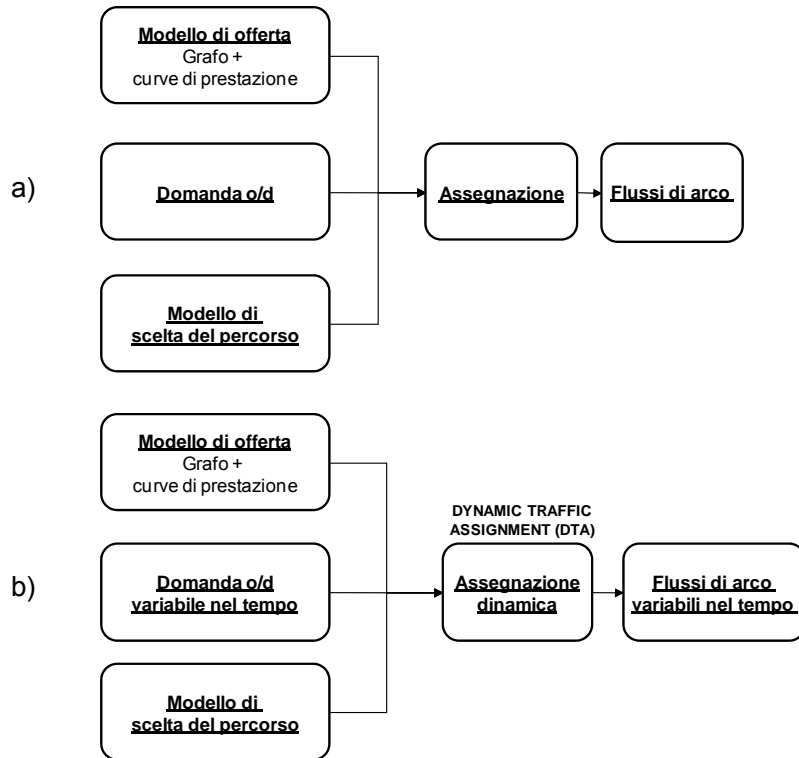


Figura 2.2: Simulazione statica e dinamica

I modelli di assegnazione dinamica (DTA, Dynamic Traffic Assignment), presupposto teorico da cui deriva la simulazione dinamica, permettono di simulare i fenomeni di congestione (formazione e movimento delle code, colli di bottiglia, “spill-back”), diventando così strumenti fondamentali per la progettazione e l’esercizio di misure di controllo dei flussi (controllo semaforico, aumento di capacità) e della domanda (limitazioni orarie della sosta, sfasamento orari inizio/fine attività), nonché per la verifica degli effetti sulla rete di sistemi e strategie di informazione agli utenti.

Questi modelli si basano su due tipologie di approccio: un approccio in simulazione ed un approccio di tipo analitico.

Nell’approccio analitico si ricerca un’approssimazione del problema di assegnazione dinamica (User Equilibrium o System Optimum) attraverso

una formulazione matematica a vincoli espliciti. La risoluzione del problema è ottenuta con algoritmi di ottimizzazione non lineare. Vengono ulteriormente classificati all'interno dell'approccio analitico le seguenti formulazioni: programmazione matematica, controllo ottimo e disuguaglianze variazionali.

Programmazione matematica (Merchant e Nemhouse, 1978; Carey, 1987)

I modelli di assegnazione dinamica basati sulla programmazione matematica formulano il problema attraverso una discretizzazione temporale delle variabili in gioco. Si mostrano però carenti nella rappresentazione della realtà e dei comportamenti degli utenti (nel System Optimum accade che i veicoli possano “scavalcare” i veicoli che li precedono – violazione della regola FIFO – o che vengano sottoposti ad attese ai nodi notevoli pur di minimizzare i costi complessivi della rete). Per limitare tali carenze è necessario aggiungere dei vincoli espliciti che rendono il problema non convesso e che ne aumentano la complessità, soprattutto per applicazioni in tempo reale su reti di ampia scala.

Controllo ottimo (Friesz et al., 1989; Wie, 1991)

A differenza della programmazione matematica, nel caso di controllo ottimo la domanda ed i flussi d'arco, così come i vincoli stessi del problema, sono considerate funzioni continue del tempo. I modelli di assegnazione dinamica basati sul controllo ottimo si dimostrano in grado di descrivere adeguatamente la dinamicità del sistema; tuttavia hanno differenti limitazioni (perdita di vincoli espliciti per rappresentare la regola FIFO o per evitare eccessive attese ai nodi) che determinano una non adeguata e spesso non realistica rappresentazione della congestione. Problema ancor più determinante per tale approccio è la mancanza di una procedura risolutiva generale che possa essere applicata su qualsivoglia rete.

Disuguaglianze variazionali (Friesz, 1993; Ran e Boyce, 1996; Chen e Hsueh, 1998)

L'approccio basato sulle disuguaglianze variazionali presenta una serie di vantaggi quali una maggiore generalità della formulazione e dunque una maggiore semplicità di dimostrare proprietà matematiche fondamentali, nonché di permettere eventuali estensioni del problema e di effettuare analisi di sensitività. Tuttavia le comuni limitazioni degli approcci

precedenti rimangono e qualora si ricerchi una maggiore congruenza con la rappresentazione di fenomeni reali, la complessità computazionale aumenta portando ad una difficoltà di applicazione su reti di dimensioni reali.

Nell'approccio in simulazione viene utilizzato un simulatore di traffico per riprodurre le complesse dinamiche del traffico veicolare. In particolare il simulatore riesce a tenere conto dei complessi vincoli che regolano la propagazione del flusso sulla rete e delle interazioni spazio-tempo (incidenza archi-percorsi, conservazione del flusso, manovre veicolari).

Un'adeguata trattazione degli approcci analitici ed in simulazione presenti in letteratura può essere trovata in Peeta e Ziliaskopoulos (2001).

2.2 La stima della domanda di trasporto

La domanda di trasporto risulta di fondamentale importanza nei modelli di simulazione del traffico. Difatti sia l'analisi che la progettazione di interventi sui sistemi di trasporto richiedono la conoscenza della domanda attuale e/o la previsione di quella futura: solo dalla interazione domanda-offerta è possibile definire, tramite il processo di assegnazione, le caratteristiche del deflusso (velocità, densità e flusso) che interessano i diversi elementi infrastrutturali presenti nel sistema di analisi, quindi valutare il livello di servizio ed i costi totali del trasporto necessari per la fase progettuale.

Ovviamente, essendo la domanda futura una proiezione attraverso modelli matematici della domanda attuale, maggiore sarà la bontà di quest'ultima e maggiore sarà la bontà della valutazione del sistema nello stato futuro. Risulta dunque necessario partire sempre da una buona stima di quella che è la matrice di domanda all'orizzonte temporale attuale.

Per ottenere la domanda allo stato attuale è possibile procedere tramite una stima diretta o tramite una stima da modello.

Avere una stima diretta della domanda attuale non è un'operazione sempre praticabile in quanto sono necessarie informazioni sulle caratteristiche degli spostamenti effettuati da tutti gli utenti nel periodo di riferimento e ciò può essere determinato solo tramite censimento. Per questo si procede solitamente ad una stima di tipo campionario, dove le informazioni necessarie vengono reperite attraverso il campionamento di un sottoinsieme di utenti del sistema di studio; dal campionamento è

possibile poi determinare la domanda tramite le tecniche della statistica inferenziale.

In alternativa la domanda attuale può essere stimata dalla conoscenza di pregresse informazioni censuarie e dal loro aggiornamento attraverso dati sperimentali tramite l'utilizzo di un sistema di modelli matematici.

In quest'ultimo caso si richiede che i modelli utilizzati vengano specificati, calibrati e validati, cosa che può essere fatta tramite la conoscenza o di informazioni disaggregate relative ad un campione di individui o di dati di tipo aggregato.

Negli anni ha sempre più preso piede l'utilizzo di misure aggregate di traffico ed in particolare di conteggi veicolari in quanto risultano informazioni ottenibili facilmente ed a basso costo, mentre le indagini campionarie sono di contro complesse e costose.

2.2.1 La stima diretta della domanda di trasporto attuale

La conoscenza esatta della domanda richiede informazioni sulle caratteristiche degli spostamenti di tutti gli utenti della rete nel periodo di riferimento esaminato, ovvero necessita di informazioni di tipo censuario. I costi di un rilevamento di tal genere sono però impraticabili e per questo, a meno che il dato censuario non sia già a disposizione, tutte le stime dirette di domanda sono di tipo campionario.

Alcune tipologie di indagine campionaria sono le indagini durante il viaggio o "*a bordo*", nelle quali si intervista un campione di utenti per uno o più modi di trasporto, a bordo strada nel caso di autovettura o sui mezzi ed ai terminali nel caso di utenti del trasporto pubblico. Esistono poi le indagini *a domicilio* durante le quali vengono intervistate un sottoinsieme di persone residenti all'interno dell'area di studio (tramite intervistatore che si reca direttamente presso la residenza o tramite interviste telefoniche). Ulteriori indagini sono quelle *a destinazione* (posti di lavoro, scuole, negozi, ecc.) e quelle *postali* nelle quali gli utenti vengono intervistati per posta.

Attraverso la teoria del campionamento è possibile pianificare le varie fasi dell'indagine ed in particolare:

1. la definizione dell'unità di campionamento (persona, famiglia, veicolo ecc.) ed il metodo di conteggio dell'universo;

2. la strategia di campionamento con la quale estrarre il campione di individui da intervistare;
3. l'estimatore da adottare con il quale ottenere la stima della quantità incognita;
4. la numerosità del campione da estrarre.

Nel caso di indagini campionarie, il numero di interviste necessarie a raccogliere le dovute informazioni è molto più elevato di quelle necessarie a calibrare i modelli di domanda. Inoltre raramente si ricorre nelle applicazioni ad un solo tipo di indagine campionaria, in quanto con ciascuna di queste presa singolarmente sarebbe possibile cogliere solo alcune delle componenti della domanda di trasporto.

2.2.2 La stima disaggregata dei modelli di domanda

Nel caso si ricorra a modelli matematici per la stima della domanda di trasporto, tali modelli devono essere come già detto specificati, calibrati e validati. Per procedere in tali operazioni è necessario ricorrere a delle indagini su di un campione di utenti con il fine di ottenere un determinato numero di informazioni.

Le indagini utilizzate possono essere relative a comportamenti effettivamente rilevati o dimostrati dagli utenti in un contesto reale (indagine sulle preferenze rilevate – Revealed Preferences RP) o indagini relative ai comportamenti dichiarati dagli utenti in contesti ipotetici (indagini sulle preferenze dichiarate – Stated Preferences SP).

Una volta specificato il tipo di modello, ovvero specificata la forma funzionale e gli attributi che lo compongono, si passa alla fase di calibrazione dove i singoli parametri vengono ad essere individuati dal punto di vista numerico. Il metodo più utilizzato nella fase di calibrazione di un modello di domanda di tipo disaggregato è il metodo della massima verosimiglianza (Maximum Likelihood ML) che fornisce i valori dei parametri incogniti che massimizzano la probabilità di osservare le scelte di un campione di utenti.

La funzione di verosimiglianza è espressa, nel caso di campionamento casuale semplice ovvero di interviste indipendenti, come:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1, \dots, n} p^i [j(i)](\mathbf{x}^i, \boldsymbol{\beta}, \boldsymbol{\theta}) \quad (2.2)$$

dove:

β, θ = parametri del modello

X^i = attributi di scelta relativi all'individuo i

$j(i)$ = alternativa effettivamente scelta dall'individuo i

Da cui ne deriva che la stima di massima verosimiglianza dei vettori di parametri β e θ è ottenuta massimizzando la (2.2).

2.3 Problema di correzione della matrice O/D tramite conteggi di traffico (caso statico)

A partire dalla conoscenza di una matrice originaria, è possibile aggiornare tale informazione all'attualità anche attraverso dati di tipo aggregato come i conteggi di traffico, utilizzando una procedura ricorsiva che mette a confronto i valori dei flussi derivanti dall'assegnazione della domanda di trasporto con i flussi misurati tramite conteggio sugli archi stradali della rete. La matrice finale di questa procedura dovrà produrre una buona corrispondenza tra flussi simulati e flussi rilevati, tale da garantire in fase previsionale un'adeguata valutazione dello stato del sistema.

L'importanza dei conteggi come dato aggregato da utilizzare per la stima della domanda di trasporto si evince sia dalle numerose imprecisioni che possono essere legate alla stima diretta o alla stima da modello, sia dai costi e dalla complessità connessa alle indagini campionarie. I conteggi di traffico sono economici, possono essere ottenuti in modo automatico e possono essere limitati ad un numero discreto di elementi della rete.

Il problema di correzione della matrice O/D tramite conteggi di traffico può essere visto come il problema inverso al problema di assegnazione (Figura 2.3).

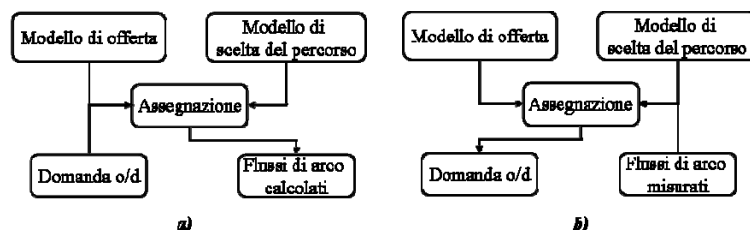


Figura 2.3: Modello di assegnazione (a) e problema di stima della domanda di trasporto (b)

Di fatti mentre nell'assegnazione, partendo dalla conoscenza dell'offerta, della domanda e del modello che regola la scelta del percorso, si arriva a determinare il valore dei flussi di arco sulla rete, nella stima della matrice O/D si parte dai flussi di arco misurati, dalla conoscenza dell'offerta e del modello di scelta del percorso e si arriva alla definizione della domanda.

Tale problema è stato formalizzato da Cascetta e Nguyen nel 1987 come un problema di ottimizzazione del tipo:

$$\mathbf{d}^* = \arg \min_{\mathbf{x} \geq 0} [f_1(\mathbf{x}, \hat{\mathbf{d}}) + f_2(\mathbf{v}(\mathbf{x}), \hat{\mathbf{v}})] \quad (2.3)$$

dove

$$\hat{\mathbf{v}} = \mathbf{v} + \boldsymbol{\varepsilon}^{OBS} = \hat{\mathbf{M}}\mathbf{x} + \boldsymbol{\varepsilon}^{SIM} + \boldsymbol{\varepsilon}^{OBS} = \mathbf{v}(\mathbf{x}) + \boldsymbol{\varepsilon} \quad (2.4)$$

con

$\hat{\mathbf{d}}$ = matrice di domanda "target "

\mathbf{x} = matrice di domanda stimata

$\hat{\mathbf{M}}$ = matrice di assegnazione derivante da modello

$\hat{\mathbf{v}}$ = flussi d'arco derivati da conteggi di traffico

\mathbf{v} = flussi d'arco derivati da assegnazione della domanda \mathbf{x}

$\boldsymbol{\varepsilon}^{OBS}, \boldsymbol{\varepsilon}^{SIM}$ = errori di osservazione e di simulazione

In pratica si ricerca quella matrice \mathbf{d}^* che minimizza una determinata misura della distanza tra domanda stimata \mathbf{x} e domanda iniziale nota (matrice target $\hat{\mathbf{d}}$), nonché della distanza tra flussi d'arco derivanti dall'assegnazione della matrice stimata $\mathbf{v}(\mathbf{x})$ e conteggi di traffico $\hat{\mathbf{v}}$.

Questa misura di distanza viene fornita tramite le funzioni f_1, f_2 di *goodness of fit*, le quali dipendono dal tipo di informazioni disponibili (sperimentali o non sperimentali) e dalle leggi di probabilità associate alle informazioni stesse.

La presenza di una matrice di domanda target nella funzione obiettivo diventa essenziale per la indeterminatezza che caratterizza il problema attraverso l'uso dei soli conteggi di traffico. Difatti il numero di coppie OD è usualmente molto più grande del numero di conteggi disponibili e le

matrici di domanda corrispondenti ad un determinato vettore di conteggi possono così essere infinite (Figura 2.4).

Per limitare questa indeterminazione si utilizza per l'appunto una matrice target e si seleziona tra l'infinito numero di possibili soluzioni candidate, quella che le si avvicina di più. In questo modo il problema ha un'unica soluzione (Bierlaire, 2002).

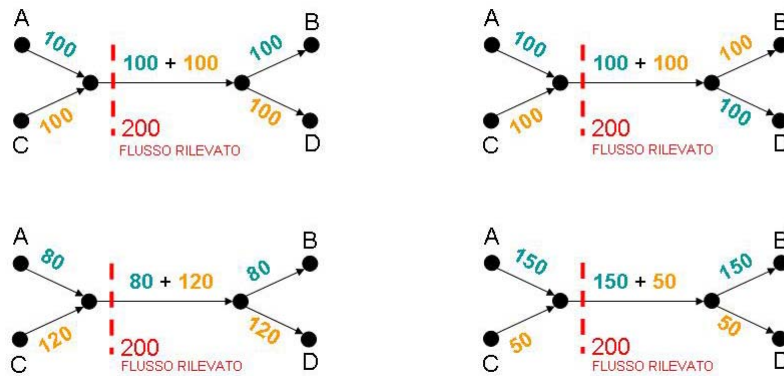


Figura 2.4: Indeterminatezza del problema di stima della domanda da conteggi di traffico

Il problema di stima della matrice O/D riportato in (2.3) può essere formulato, nel caso di reti congestionate, come problema di punto fisso; difatti considerando la dipendenza tra costi e flussi d'arco $\mathbf{c} = \mathbf{c}(\mathbf{v})$ e tra flussi d'arco e domanda di trasporto $\mathbf{v}(\mathbf{x}) = \hat{\mathbf{M}}\mathbf{x}$ attraverso la matrice di assegnazione, si ottiene:

$$\mathbf{v}(\mathbf{x}) = \hat{\mathbf{M}}\mathbf{x} = \hat{\mathbf{M}}(\mathbf{c})\mathbf{x} = \hat{\mathbf{M}}(\mathbf{c}(\mathbf{v}))\mathbf{x} = \hat{\mathbf{M}}(\mathbf{c}(\mathbf{v}(\mathbf{d})))\mathbf{x} \quad (2.5)$$

In alternativa il problema è esprimibile come un problema di ottimizzazione bilivello del tipo:

$$\begin{aligned} \mathbf{d}^* &= \arg \min_{\mathbf{x} \geq 0} [f_1(\mathbf{x}, \hat{\mathbf{d}}) + f_2(\mathbf{v}(\mathbf{x}), \hat{\mathbf{v}})] \\ \mathbf{v}(\mathbf{x}) &= \arg \min f(\hat{\mathbf{v}}) \end{aligned} \quad (2.6)$$

Nel caso di esplicitazione del problema di correzione della matrice O/D come problema di punto fisso, è possibile adottare una procedura risolutiva di tipo iterativo che risolve alternativamente il problema della

stima e quello dell'assegnazione mediando i risultati ottenuti sino a convergenza.

In Figura 2.5 viene riportato a titolo di esempio la procedura da seguire in caso di utilizzo del metodo iterativo delle medie successive (MSA).

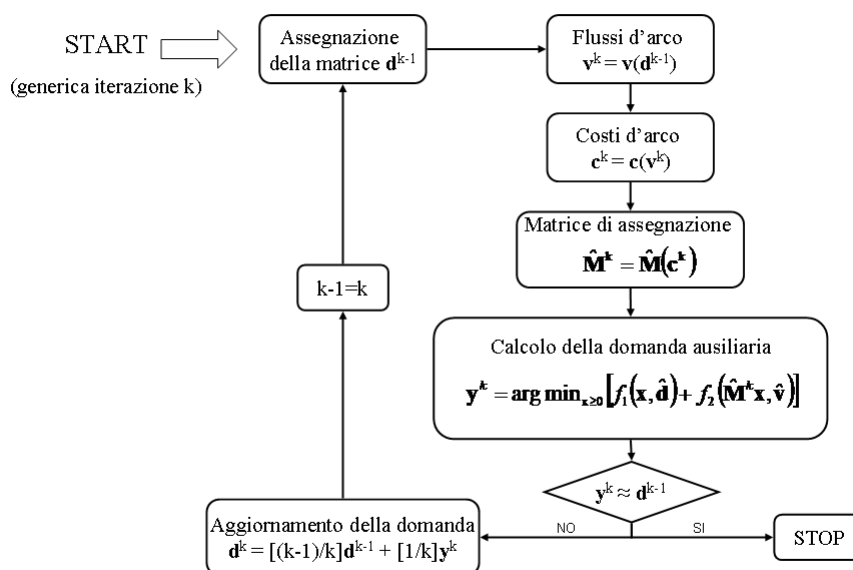


Figura 2.5: MSA per la risoluzione del problema di stima O/D

2.3.1 Stimatori o misure di goodness of fit

Le due funzioni f_1, f_2 presenti nella (2.6) sono i cosiddetti indicatori di “bontà” (*goodness of fit*) e la loro espressione dipende dalle leggi di probabilità utilizzate per descrivere le informazioni sperimentali e non sperimentali disponibili e richieste dal modello (stima a priori della domanda, flussi derivanti da conteggi, flussi derivanti da simulazione).

Per esplicitare le due funzioni si ricorre alla teoria degli stimatori, la quale comprende:

1. la teoria della massima verosimiglianza (ML);
2. la teoria dei minimi quadrati generalizzati (GLS);
3. i metodi Bayesiani.

Secondo la teoria della massima verosimiglianza la domanda da stimare è frutto della massimizzazione della probabilità di osservare i risultati delle indagini campionarie e dei conteggi di traffico:

$$\mathbf{d}^{ML} = \arg \max_{\mathbf{d} \in S} [\ln L(\mathbf{n}|\mathbf{x}) + \ln L(\hat{\mathbf{v}}|\mathbf{x})] \quad (2.7)$$

con:

\mathbf{n} = vettore risultati indagini campionarie sulla domanda

\mathbf{x} = vettore di domanda incognito

$\hat{\mathbf{v}}$ = conteggi di traffico

$\ln L(\mathbf{n} | \mathbf{x})$ = logaritmo della probabilità di osservare il vettore campionario \mathbf{n} se \mathbf{x} è la domanda reale

$\ln L(\hat{\mathbf{v}} | \mathbf{x})$ = logaritmo della probabilità di osservare i conteggi $\hat{\mathbf{v}}$ se \mathbf{x} è la domanda reale

Per risolvere il problema (2.7) è necessario formulare delle ipotesi sulle leggi di probabilità condizionate al vettore di domanda \mathbf{x} dei conteggi di domanda \mathbf{n} e di traffico $\hat{\mathbf{v}}$.

Per i conteggi di traffico si assume che questi siano variabili aleatorie del tipo $\hat{\mathbf{v}} = \mathbf{v}(\mathbf{x}) + \boldsymbol{\varepsilon}$ esprimibili dunque tramite un valor medio ed il rispettivo residuo aleatorio e che i residui aleatori siano indipendenti e distribuiti secondo una distribuzione di Poisson o che seguano una distribuzione del tipo multivariata normale (MVN).

Per i conteggi di domanda \mathbf{n} , la funzione di massima verosimiglianza dipende dalle modalità con cui è stato ottenuto il campione:

1. nel caso di campionamento casuale per zona di origine, l'ipotesi è di variabile multinomiale;
2. nel caso di campionamento casuale per zona di origine (>10 per ogni origine), si ha il prodotto di variabili di Poisson.

Secondo la teoria dei Minimi Quadrati Generalizzati (Generalized Least Squares – GLS) è possibile ottenere la stima del vettore dei flussi di domanda a partire da un sistema di equazioni stocastiche, lineari, nel vettore dei flussi di domanda stesso.

Le equazioni in gioco sono:

$$\begin{aligned} \hat{\mathbf{v}} &= \hat{\mathbf{M}}\mathbf{x} + \boldsymbol{\varepsilon} & E(\boldsymbol{\varepsilon}) &= 0, \text{Var}(\boldsymbol{\varepsilon}) = \mathbf{W} \\ \hat{\mathbf{d}} &= \mathbf{x} + \boldsymbol{\eta} & E(\boldsymbol{\eta}) &= 0, \text{Var}(\boldsymbol{\eta}) = \mathbf{Z} \end{aligned} \quad (2.8)$$

Solitamente le matrici di varianza-covarianza \mathbf{W} e \mathbf{Z} sono matrici diagonali, ovvero si trascura la covarianza, sia in quanto difficile da determinare, sia per risparmiare risorse di calcolo.

In tal caso non sono necessarie ipotesi sui residui, ma la domanda incognita deriva dalla seguente minimizzazione:

$$\mathbf{d}^{GLS} = \arg \min_{\mathbf{x} \geq 0} \left[(\hat{\mathbf{d}} - \mathbf{x})^T \mathbf{Z}^{-1} (\hat{\mathbf{d}} - \mathbf{x}) + (\hat{\mathbf{v}} - \hat{\mathbf{M}}\mathbf{x})^T \mathbf{W}^{-1} (\hat{\mathbf{v}} - \hat{\mathbf{M}}\mathbf{x}) \right] \quad (2.9)$$

dove:

$\hat{\mathbf{d}}$ = domanda da campionamento;

$\hat{\mathbf{M}}$ = matrice di assegnazione;

\mathbf{x} = vettore di domanda incognito;

$\hat{\mathbf{v}}$ = conteggi di traffico;

\mathbf{Z} = matrice var – cov della domanda da campionamento rispetto alla domanda incognita;

\mathbf{W} = matrice var – cov dei conteggi di traffico rispetto ai flussi simulati.

Infine secondo i metodi Bayesiani la domanda è stimata combinando informazioni sperimentali o campionarie con altre non sperimentali o soggettive. In particolare nel caso in esame le informazioni sperimentali sul vettore di domanda incognito sono quelle relative ai conteggi di traffico, mentre le informazioni non sperimentali possono essere relative a vecchie stime della matrice origine-destinazione da aggiornare, a stime ottenute con modelli di domanda oppure a semplici aspettative dell'analista.

In quest'ultimo caso la domanda incognita deriva dalla seguente massimizzazione:

$$\mathbf{d}^B = \arg \max_{\mathbf{x} \in S} \left[\ln g(\mathbf{x} | \hat{\mathbf{d}}) + \ln L(\hat{\mathbf{v}} | \mathbf{x}) \right] \quad (2.10)$$

in cui

L = funzione di massima verosimiglianza;

g = funzione di probabilità a priori;

\mathbf{x} = vettore di domanda incognito;

$\hat{\mathbf{v}}$ = conteggi di traffico;

$\hat{\mathbf{d}}$ = domanda nota a priori.

L'espressione delle funzioni di probabilità g ed L dipendono dalle ipotesi sulle loro distribuzioni.

Per i conteggi di traffico $\hat{\mathbf{v}} = \mathbf{v}(\mathbf{x}) + \boldsymbol{\varepsilon}$, le $\boldsymbol{\varepsilon}$ sono o indipendenti e distribuite secondo una distribuzione di Poisson o seguono una distribuzione del tipo multivariata normale (MVN).

Per le leggi di probabilità, condizionate al vettore di domanda \mathbf{d} , della matrice incognita si può ipotizzare:

1. una distribuzione multinomiale per le variabili aleatorie:
 $\mathbf{x} = E(\mathbf{x}) + \boldsymbol{\varepsilon}$;
2. una distribuzione di Poisson e l'indipendenza dei residui:
 $\mathbf{x}_{od} = \hat{\mathbf{d}}_{od} + \boldsymbol{\varepsilon}$;
3. una distribuzione multivariata normale: $\mathbf{x} = \hat{\mathbf{d}} + \boldsymbol{\varepsilon}$; $Var(\boldsymbol{\varepsilon}) = \mathbf{Z}_B$.

In Tabella 2.1 sono riportate le definizioni delle misure di *goodness of fit* presenti nella (2.6) per le differenti teorie degli stimatori e per le differenti ipotesi effettuate sui residui aleatori.

Tabella 2.1: Definizione degli stimatori per le differenti teorie adottate ed ipotesi sui residui

Distanza dalla stima iniziale $f_1(\mathbf{x}, \hat{\mathbf{d}})$	Distanza dai conteggi di traffico $f_2(\mathbf{v}(\mathbf{x}), \hat{\mathbf{v}})$
GLS $(\hat{\mathbf{d}} - \mathbf{x})^T \mathbf{Z}^{-1} (\hat{\mathbf{d}} - \mathbf{x})$	GLS $(\hat{\mathbf{v}} - \mathbf{v}(\mathbf{x}))^T \mathbf{W}^{-1} (\hat{\mathbf{v}} - \mathbf{v}(\mathbf{x}))$
ML	ML
• MVN $(\hat{\mathbf{d}} - \mathbf{x})^T \mathbf{Z}^{-1} (\hat{\mathbf{d}} - \mathbf{x})$	• MVN $(\hat{\mathbf{v}} - \mathbf{v}(\mathbf{x}))^T \mathbf{W}^{-1} (\hat{\mathbf{v}} - \mathbf{v}(\mathbf{x}))$
• Poisson $-\sum_{od} (n_{od} \ln(\alpha_{od} x_{od}) - \alpha_{od} x_{od})$	• Poisson $\sum_i (\hat{v}_i \ln v_i(\mathbf{x}) - v_i(\mathbf{x}))$
Bayes	Bayes
• MVN $(\hat{\mathbf{d}} - \mathbf{x})^T \mathbf{Z}^{-1} (\hat{\mathbf{d}} - \mathbf{x})$	• MVN $(\hat{\mathbf{v}} - \mathbf{v}(\mathbf{x}))^T \mathbf{W}^{-1} (\hat{\mathbf{v}} - \mathbf{v}(\mathbf{x}))$
• Poisson $\sum_{od} x_{od} \ln(x_{od} / \hat{d}_{od} - 1)$	• Poisson $\sum_i (\hat{v}_i \ln v_i(\mathbf{x}) - v_i(\mathbf{x}))$
• Multinomiale $\sum_{od} x_{od} \ln(x_{od} / \hat{d}_{od})$	

2.4 *Problema di correzione della matrice O/D tramite conteggi di traffico (caso dinamico)*

Nel passaggio da modelli statici a modelli dinamici, la domanda varia all'interno del periodo di riferimento secondo l'ipotesi di dinamica intraperiodale.

In quest'ultimo caso la rappresentazione matriciale vede la presenza di un numero n di matrici, tante quanti gli intervalli temporali in cui è stato suddiviso il periodo di riferimento T .

Ovviamente l'indeterminatezza associata al problema di correzione della matrice O/D va ad aumentare con l'aggiunta della dimensione temporale e di conseguenza andrà ad aumentare la complessità stessa del problema.

Inoltre la fase di carico della rete, con una domanda che segue l'ipotesi di dinamica intraperiodale, diventa maggiormente complessa in quanto gli utenti partiti in un certo intervallo temporale lungo un certo percorso contribuiranno al flusso sugli archi che compongono il percorso stesso esclusivamente negli intervalli in cui ciascun arco viene raggiunto e attraversato; inoltre in reti congestionate, il tempo necessario per raggiungere un arco dipende a sua volta dal flusso incontrato sugli archi precedenti lungo lo stesso percorso (Cascetta, 1992).

La comune formalizzazione non è altro che una estensione del caso statico, data però la presenza della dimensione temporale si distingue tra approccio simultaneo ed approccio sequenziale (Cascetta et al., 1993).

Nel caso di stima simultanea si ha:

$$(\mathbf{d}_1^* \dots \mathbf{d}_{n_h}^*) = \arg \min_{(\mathbf{x}_1 \dots \mathbf{x}_{n_h}) \geq 0} [f_1(\mathbf{x}_1 \dots \mathbf{x}_{n_h}, \hat{\mathbf{d}}_1 \dots \hat{\mathbf{d}}_{n_h}) + f_2(\mathbf{v}_1 \dots \mathbf{v}_{n_h}, \hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_{n_h})] \quad (2.11)$$

Si tratta quindi di individuare le matrici $\mathbf{d}_1^* \dots \mathbf{d}_{n_h}^*$, per ogni intervallo temporale n_h in cui è suddiviso il periodo di analisi T , che minimizzano una determinata misura della distanza tra domanda stimata $\mathbf{x}_1^* \dots \mathbf{x}_{n_h}^*$ e domanda iniziale nota (matrici target $\hat{\mathbf{d}}_1 \dots \hat{\mathbf{d}}_{n_h}$), nonché della distanza tra flussi d'arco derivanti dall'assegnazione della matrice stimata $\mathbf{v}_1 \dots \mathbf{v}_{n_h}$ e conteggi di traffico $\hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_{n_h}$.

Per la stima sequenziale:

$$\mathbf{d}_{n_h}^* = \arg \min_{\mathbf{x}_{n_h} \geq 0} [f_1(\mathbf{x}_{n_h}, \hat{\mathbf{d}}_{n_h}) + f_2(\mathbf{v}_{n_h}(\mathbf{x}_{n_h} | \mathbf{d}_1^* \dots \mathbf{d}_{n_h-1}^*), \hat{\mathbf{v}}_{n_h})] \quad (2.12)$$

Viene così ad essere individuata una matrice alla volta, partendo dal primo intervallo temporale e procedendo sino all'ultimo, mantenendo di volta in volta fisse le matrici precedentemente calcolate.

In quest'ultimo caso si ottiene una riduzione della complessità computazionale del problema in quanto il problema stesso viene ad essere scomposto in sottoproblemi più semplici da trattare ed ogni matrice stimata può essere utilizzata come stima iniziale per l'intervallo temporale successivo.

Ulteriori approcci presenti sono:

- approccio di Ashok;
- approccio di Van der Zijpp;
- approccio di Balakrishna.

2.4.1 Approccio di Ashok

Nel 1996 Ashok sviluppa una stima sequenziale della matrice O/D nel caso dinamico che utilizza una equazione di transizione per catturare l'evoluzione dello stato del sistema ed una equazione di misura per incorporare le misure derivanti dai sensori di traffico. La maggiore novità è in questo caso la definizione di stato del sistema in termini di "deviazione", ovvero in termini di differenza tra flussi OD stimati e flussi OD attesi (dati storici). Ciò permette di rappresentare lo stato del sistema attraverso una distribuzione simmetrica, la quale possiede proprietà utili al processo di stima e che non possono essere fornite direttamente dai flussi OD.

L'equazione di transizione è basata su un processo auto regressivo e descrive l'evoluzione del sistema da un intervallo all'altro:

$$\mathbf{x}_{h+1} - \mathbf{x}_{h+1}^H = \sum_{p=h-q'+1}^h \mathbf{f}_{h+1}^p (\mathbf{x}_p - \mathbf{x}_p^H) + \mathbf{w}_{h+1} \quad (2.13)$$

dove:

\mathbf{x}_{h+1} è la domanda che parte nell'intervallo $h+1$

\mathbf{x}_{h+1}^H è il migliore dato storico relativo alla domanda in partenza nell'intervallo $h+1$

\mathbf{f}_{h+1}^p è una matrice relativa alle relazioni spaziali e temporali intercorrenti tra gli intervalli p e $h+1$. Il suo calcolo comprende la stima di una serie di modelli di regressione lineare per ogni coppia O-D.

q' è il grado del processo di auto regressione e rappresenta la lunghezza nel tempo della storia passata che influenza l'intervallo corrente.

L'equazione di misura cattura invece la relazione tra domanda e flussi d'arco sempre in termini di deviazione rispetto al dato storico ed è espressa come:

$$\mathbf{y}_h - \mathbf{y}_h^H = \sum_{p=h-p'}^h \mathbf{a}_h^p (\mathbf{x}_p - \mathbf{x}_p^H) + \mathbf{v}_h \quad (2.14)$$

con

$$\mathbf{y}_h^H = \sum_{p=h-p'}^h \mathbf{a}_h^p \mathbf{x}_p^H \text{ che fornisce una stima dello storico dei conteggi } \mathbf{y}_h^H ;$$

\mathbf{a}_h^p è la matrice di assegnazione che relaziona la domanda in partenza nell'intervallo p ai flussi d'arco dell'intervallo h ;

\mathbf{v}_h è la matrice degli errori di misura.

L'autore utilizza il filtro di Kalman come metodo risolutivo per la stima di \mathbf{x}_h : nel primo step i flussi \mathbf{x}_h vengono stimati sequenzialmente per ogni intervallo h (ignorando il contributo di \mathbf{x}_h sui conteggi $\mathbf{y}_{h+1}, \mathbf{y}_{h+2}, \dots, \mathbf{y}_H$). Ogni matrice OD viene quindi aggiornata partendo dall'ultimo intervallo H e risalendo sino al primo ($h=1$). L'informazione contenuta in $\mathbf{y}_{h+1}, \mathbf{y}_{h+2}, \dots, \mathbf{y}_H$ viene così utilizzata completamente nell'identificare i flussi per l'intervallo h .

2.4.2 Approccio di Van der Zijpp

Van der Zijpp (1996) sviluppa un metodo per la stima della matrice OD dinamica nel caso di reti autostradali in cui i confini temporali del problema sono determinati analizzando le traiettorie dei veicoli. Assumendo note le velocità dei veicoli, vengono tracciate le traiettorie del primo e dell'ultimo veicolo per ogni intervallo temporale. Le traiettorie di tutti i rimanenti veicoli sono quindi determinate secondo una legge di tipo first in – first out (FIFO). Il set di traiettorie viene utilizzato per determinare le quote di domanda da allocare al singolo sensore di conteggio. Tali quote (“*split fractions*”) vengono modellate secondo una distribuzione del tipo multivariata normale troncata (*Truncated*

Multivariate Normal Distribution - TMVN) e sono aggiornate ad ogni step temporale tramite la formula di Bayes.

Il problema in questo caso sta nel calcolo del valor medio della distribuzione multivariata normale troncata, per la quale non esiste una soluzione in forma chiusa; ulteriore problema è la mancata esistenza di una espressione analitica per il calcolo della precedente distribuzione. Infine il calcolo delle traiettorie dei veicoli richiede la conoscenza delle velocità lungo tutto l'itinerario, per ogni intervallo temporale, per cui la velocità assume un ruolo fondamentale nella formulazione del modello.

2.4.3 Approccio di Balakrishna

Nel 2006 Balakrishna propone un approccio di stima simultaneo della matrice OD applicato al caso dinamico caratterizzato dai seguenti aspetti:

1. all'interno del problema di ottimizzazione vengono inglobate differenti tipologie di misure nonché i parametri di scelta del percorso e dell'offerta;
2. la formulazione non dipende dalla matrice di assegnazione.

Non vengono più ad essere i conteggi di traffico dunque le sole informazioni utilizzate durante la procedura di stima, ma si può far riferimento a velocità e densità, nonché a misure dei tempi di percorrenza disponibili tramite sistemi AVI (Automatic Vehicle Identification) o veicoli sonda. La dipendenza tra queste misure ed i flussi OD non viene catturata più tramite la matrice di assegnazione, ma utilizzando il modello di assegnazione come una scatola nera ovvero direttamente in simulazione.

Il problema è formulato secondo il seguente schema di ottimizzazione:

$$\text{Min}_{\mathbf{x}_1, \dots, \mathbf{x}_H} \sum_{h=1}^H [z_1(\mathbf{M}_h, \mathbf{N}_h) + z_2(\mathbf{x}_h, \mathbf{x}_h^a)] \quad (2.15)$$

soggetto ai seguenti vincoli:

$$\mathbf{N}_h = f(\mathbf{x}_1, \dots, \mathbf{x}_H; \boldsymbol{\beta}; G_1, \dots, G_h) \quad \forall h \in \{1, 2, \dots, H\} \quad (2.16)$$

dove:

\mathbf{M}_h e \mathbf{N}_h sono rispettivamente le misure osservate e simulate per l'intervallo temporale h ;

\mathbf{x}_h è il vettore dei flussi OD che partono nell'intervallo h ;

\mathbf{x}_h^a è il corrispondente vettore noto a priori dei flussi OD che partono nell'intervallo h ;

$f()$ fornisce gli output del modello di assegnazione dinamica in funzione delle matrici OD, della rete G nell'intervallo h e dei parametri β del modello.

L'algoritmo risolutivo utilizzato è in questo caso dato da una combinazione del metodo di Box-Complex (Box, 1965) e di SNOBFIT (Stable Noisy Optimization by Branch and Fit, Huyer e Neumaier, 2004). Il primo viene utilizzato per generare un set di soluzioni iniziali nei pressi di potenziali minimi locali, il secondo utilizza un'approssimazione quadratica per migliorare la ricerca locale a partire da tali punti.

Tale approccio è stato applicato su una porzione della città di Los Angeles in California con buoni miglioramenti della corrispondenza di flussi e velocità (miglioramento di circa il 50% in termini di RMSN).

2.5 Osservazioni

Effettuando una valutazione degli approcci presenti in letteratura per la stima della matrice OD nel caso dinamico, si può senz'altro affermare che l'approccio simultaneo risulta formalmente più corretto dell'approccio sequenziale; infatti in quest'ultimo si riduce la complessità computazionale del problema, ma i conteggi degli intervalli seguenti al generico intervallo h , non vengono utilizzati per aggiornare le matrici di domanda degli intervalli precedenti ad h : tale aspetto, nel caso di reti congestionate di elevate dimensioni con spostamenti che coprono diversi intervalli temporali, può portare ad una stima non soddisfacente della matrice o/d dinamica.

Per ottenere una buona stima è necessario dunque puntare su un approccio simultaneo.

La relazione tra flussi d'arco e flussi di domanda è esplicitata in genere tramite la matrice di assegnazione (2.4); in realtà nel caso dinamico si ha a che fare con un insieme di matrici, tante quante gli intervalli temporali in cui è suddiviso il periodo di riferimento:

$$\hat{\mathbf{v}}_h = \sum_{p=h-p'}^h \hat{\mathbf{M}}_h^p \mathbf{x}_p + \boldsymbol{\varepsilon}_h \quad (2.17)$$

con

-
- $\hat{\mathbf{v}}_h$ = flussi d'arco nell'intervallo h derivanti da conteggi di traffico;
 - $\hat{\mathbf{M}}_h^p$ = matrice di assegnazione derivante da modello;
 - \mathbf{x}_p = matrice di domanda relativa all'intervallo p ;
 - p = numero di intervalli temporali ricoperti dal percorso più lungo;
 - $\boldsymbol{\varepsilon}_h$ = errori di osservazione e di simulazione relativi all'intervallo h .

L'approccio simultaneo basato sulle matrici di assegnazione risulta avere una complessità computazionale proibitiva anche su reti di medie dimensioni (Cascetta e Russo, 1997; Toledo et al., 2003; Bierlaire e Crittin, 2004), da cui se ne deduce che non è possibile utilizzare le matrici di assegnazione per esprimere il legame flussi – domanda nel caso si voglia mantenere un approccio simultaneo.

Ulteriore considerazione è relativa alle informazioni utilizzate per la correzione della matrice: si parte generalmente da una matrice target nota e da volumi di traffico misurati, ma se queste informazioni possono essere sufficienti nel caso statico, non lo sono più quando si considera la dinamica.

Nei modelli di assegnazione statica si considera la domanda uniformemente distribuita, ipotesi strutturalmente incompatibile con i fenomeni di sovra saturazione che si verificano nella quotidianità dei contesti urbani. Per la struttura stessa del modello di offerta nel caso statico, un superamento della capacità per un determinato periodo produce un ritardo infinito; a ciò si ovvia modificando le curve di deflusso in modo tale da far sì che il ritardo calcolato possa essere finito anche per rapporti flusso/capacità superiori ad uno. Ciò comporta un valore del ritardo che non sempre riesce a rappresentare le condizioni effettive del sistema. Inoltre l'ipotesi di domanda uniforme non permette di seguire l'evolversi della congestione sulla rete nel tempo.

Per superare tali limiti nascono i modelli di assegnazione dinamica, i quali permettono di simulare esplicitamente la formazione ed il dissipamento delle code, i colli di bottiglia ed i fenomeni di spill-back. Assumono in tale contesto un'importanza rilevante grandezze quali densità e velocità, le sole che possono discriminare tra uno stato ipercritico o ipocritico.

Di fatti il flusso da solo non è in grado di dare informazioni sullo stato del sistema (Figura 2.6).

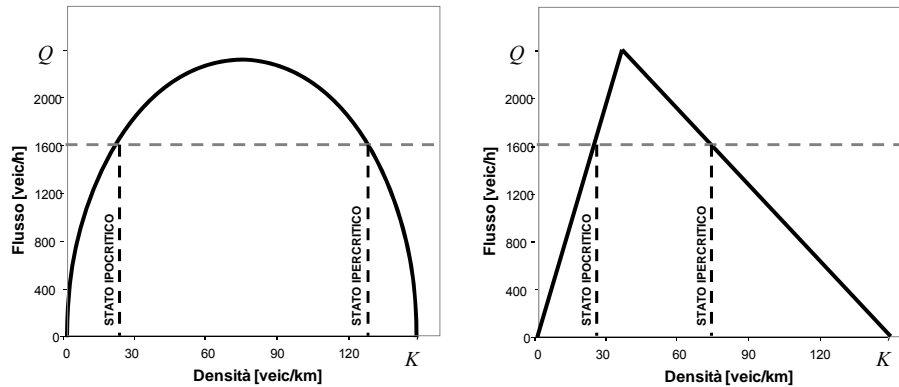


Figura 2.6: Diagramma fondamentale parabolico e triangolare ed individuazione dello stato di congestione

Proprio per questo motivo i volumi di traffico non possono essere le uniche misure utilizzate per la stima della matrice OD nel caso dinamico: è necessario acquisire ulteriori informazioni che siano in grado di definire lo stato dei singoli archi della rete e quindi ottenere una giusta stima della domanda che genera l'effettivo livello di congestione della rete.

Per far ciò si può ricorrere, proprio come nell'approccio utilizzato da Balakrishna, ad informazioni circa la velocità mantenuta dai veicoli sugli archi della rete, informazione questa facilmente reperibile attraverso le stesse apparecchiature utilizzate per i conteggi veicolari. Ulteriori informazioni circa lo stato "dinamico" del sistema possono essere ottenute attraverso i tempi di percorrenza; questi ultimi possono essere desunti tramite l'utilizzo di veicoli sonda dotati di sistemi di rilevamento come gli RFID o sistemi GPS (Eisenman e List, 2004, Zhou e Mahmassani, 2006).

Per tutti gli approcci menzionati, il miglioramento ottenibile sulla domanda attraverso la procedura di correzione dipende dalla qualità della matrice di partenza, ovvero della matrice target utilizzata. Se non è nota la bontà della matrice di partenza si può puntare verso procedure di tipo random search, così che la soluzione non rimanga confinata nei dintorni della matrice target. Kattan e Abdulhai (2006) hanno utilizzato a tal fine un Evolutionary Algorithm (EA) implementato su più calcolatori in parallelo. Chiaramente per poter adottare queste metodologie nel caso di applicazioni a contesti reali sono richieste ingenti risorse di calcolo.

3. Formalizzazione del problema

3.1 *Obiettivi*

L'obiettivo della ricerca è quello di risolvere il problema di correzione/stima della domanda di trasporto attuale in dinamica intraperiodale per un contesto di tipo off-line; non si richiede dunque l'aggiornamento in tempo reale dei valori dei flussi OD tramite le misure di traffico disponibili mano a mano che i sistemi di rilevamento acquisiscono il dato. Nel caso off-line la corretta valutazione della domanda di trasporto assume una fondamentale importanza per la successiva fase pianificatoria.

Si ipotizza di partire da una serie di matrici target note, una per ogni intervallo temporale che costituisce il periodo di riferimento e di correggere tali matrici attraverso l'utilizzo di misure di traffico.

Le matrici target derivano originariamente da informazioni censuarie e/o da modello, eventualmente ricostruite in statica attraverso la sola informazione dei flussi d'arco; si passa poi ad una suddivisione della matrice statica nei diversi intervalli temporali in funzione dell'andamento temporale delle misure di traffico a disposizione.

Le principali pecche di tale domanda di partenza sono da ricondurre al processo di correzione in statica che tende a sovrastimare i valori dei flussi OD, non tenendo conto della capacità della rete, ed alla suddivisione in intervalli temporali che presuppone la conoscenza degli orari di inizio degli spostamenti. Comunque, anche in virtù del fatto che nell'usuale pratica ingegneristica la domanda di trasporto è il presupposto fondamentale per la simulazione e la successiva attività di valutazione (e quindi per qualsiasi studio di traffico), è sempre possibile ipotizzare di partire da una precedente informazione sulla domanda sufficientemente buona.

Il problema riveste in tal senso l'ambito di una ricerca locale, per cui anche la scelta del metodo risolutivo adottato deriva dalle ipotesi effettuate circa la bontà del punto di partenza (ovvero dalla bontà delle matrici target).

La soluzione a cui si vuole giungere è una soluzione affidabile, ovvero si vuole dimostrare che la domanda stimata non comporta solo una buona corrispondenza con le misure di traffico, ma che sia effettivamente

rappresentativa degli spostamenti effettuati dagli utenti della rete nei diversi intervalli temporali. Per far ciò è necessario puntare su un approccio di tipo simultaneo che possa cogliere l'influenza di ogni matrice sui diversi intervalli temporali e considerare ulteriori elementi oltre i flussi che possano rappresentare al meglio le effettive condizioni di congestione della rete.

La procedura di risoluzione del problema di correzione/stima combinata della domanda in dinamica intraperiodale dovrà avere, una volta implementata, tempi computazionali che si sposino con l'obiettivo di individuazione della soluzione in un contesto di tipo off-line e che le permettano di poter essere utilizzata all'interno dei moderni modelli di assegnazione dinamica e dunque nella comune pratica ingegneristica.

3.2 Funzione obiettivo e vincoli

Il problema di correzione della matrice O/D viene di seguito formulato come un problema di ottimizzazione volto a minimizzare l'errore tra osservazioni (flussi, velocità) attuali e valori stimati, nonché tra domanda target e domanda stimata in un contesto dinamico.

Data

una rete $B = [N, A]$, dove:

N = nodi;

A = archi direzionali;

n_{od} = numero di coppie od;

il periodo di analisi T , diviso in n_h intervalli

un sottoinsieme di archi $S = \{1..n_s\} \in A$ forniti di sensori

Il problema è il seguente:

$$(\mathbf{d}_1^* .. \mathbf{d}_{n_h}^*) = \arg \min_{(\mathbf{x}_1 .. \mathbf{x}_{n_h}) \geq 0} [f_1(\mathbf{x}_1 .. \mathbf{x}_{n_h}, \hat{\mathbf{d}}_1 .. \hat{\mathbf{d}}_{n_h}) + f_2(\mathbf{v}_1 .. \mathbf{v}_{n_h}, \hat{\mathbf{v}}_1 .. \hat{\mathbf{v}}_{n_h}) + f_3(\mathbf{s}_1 .. \mathbf{s}_{n_h}, \hat{\mathbf{s}}_1 .. \hat{\mathbf{s}}_{n_h})] \quad (3.1)$$

con

$$\mathbf{v}_i = g(\mathbf{x}_1 .. \mathbf{x}_i)$$

$$\mathbf{s}_i = h(\mathbf{x}_1 .. \mathbf{x}_i)$$

dove:

\mathbf{x}_i = matrice stimata per l'intervallo di partenza $i \quad i = 1..n_h$

\mathbf{v}_i = flussi simulati sugli archi $\in \mathbf{S}$ per l'intervallo di partenza i

$i = 1..n_h$

\mathbf{s}_i = velocità simulate sugli archi $\in \mathbf{S}$ per l'intervallo di partenza i

$i = 1..n_h$

$\hat{\mathbf{d}}_i$ = matrice target per l'intervallo di partenza $i \quad i = 1..n_h$

$\hat{\mathbf{v}}_i$ = conteggi di traffico sugli archi $\in \mathbf{S}$ per l'intervallo di partenza i

$i = 1..n_h$

$\hat{\mathbf{s}}_i$ = velocità misurate sugli archi $\in \mathbf{S}$ per l'intervallo di partenza i

$i = 1..n_h$

g = mappa che lega il vettore dei flussi alla domanda di trasporto

h = mappa che lega il vettore delle velocità alla domanda di trasporto

soggetto ai seguenti vincoli:

$$\text{Limite Inferiore:} \quad (\mathbf{x}_1 \dots \mathbf{x}_{n_h}) \geq (\mathbf{x}_1^{L.B.} \dots \mathbf{x}_{n_h}^{L.B.})$$

$$\text{Limite Superiore:} \quad (\mathbf{x}_1 \dots \mathbf{x}_{n_h}) \leq (\mathbf{x}_1^{U.B.} \dots \mathbf{x}_{n_h}^{U.B.})$$

$$\text{Vincolo di generazione:} \quad \sum_{i=1}^{n_h} G_o^i \leq G_o^* \quad \forall o \in \{\text{origini}\}$$

con

G_o^* = generazione da modello della zona o.

I tre termini della funzione obiettivo rappresentano la distanza tra flussi e velocità simulate e le relative misure dirette e la distanza tra domanda target e la risultante domanda stimata.

Le funzioni f_1 , f_2 e f_3 dipendono, come già riportato, dalla tipologia di stimatore utilizzato e dalle informazioni disponibili (Cascetta et al. 1993). Nel caso di metodo dei minimi quadrati generalizzati (Generalized Least Squares framework - GLS) sono necessarie informazioni aggiuntive circa

l'accuratezza delle misure; queste informazioni possono essere incorporate attraverso l'introduzione di pesi interni nella funzione obiettivo (risultanti nelle matrici di varianza-covarianza) così che la funzione f_1 possa ad esempio assumere la seguente formulazione:

$$f_1(\mathbf{x}_1 \dots \mathbf{x}_{n_h}, \hat{\mathbf{d}}_1 \dots \hat{\mathbf{d}}_{n_h}) = (\mathbf{x} - \hat{\mathbf{d}})^T \mathbf{V}^{-1} (\mathbf{x} - \hat{\mathbf{d}}) \quad (3.2)$$

dove:

\mathbf{V} = matrice di varianza-covarianza del vettore degli errori di campionamento che influenzano la stima $\hat{\mathbf{d}}$.

Se questa informazione non è disponibile, i differenti termini della funzione obiettivo possono essere controllati attraverso l'uso di pesi esterni scalari che rappresentano il grado di confidenza dell'analista nei confronti delle misure di traffico (flussi e velocità) e delle osservazioni condotte sulla domanda (domanda target).

I vincoli di limite inferiore e superiore sono introdotti per eliminare la possibilità di ottenere soluzioni non possibili (flussi di domanda negativi) e per restringere lo spazio di ricerca.

Il vincolo di generazione è la principale novità introdotta nel presente approccio. Durante la procedura di stima è fondamentale considerare il quantitativo complessivo di domanda generata da ogni singola zona dell'area di studio. Le informazioni sulle generazioni nelle fasce temporali di punta del traffico (coincidenti con gli usuali periodi di analisi) sono facilmente reperibili o ricostruibili; inoltre i modelli di generazione sono i modelli che garantiscono la maggiore affidabilità tra i modelli dei sistemi di trasporto.

Il vincolo di generazione interviene in qualità di limite superiore principalmente per superare un problema che si viene a verificare nel processo di stima qualora si abbia a che fare con reti particolarmente congestionate: infatti se la congestione è forte, a causa dei vincoli di capacità della rete, i flussi sono costanti ed indipendenti dal livello di domanda. In tal caso, differenti valori di domanda possono generare la stessa corrispondenza in termini di flussi all'interno della funzione obiettivo.

Il vincolo di generazione permette di compensare la mancata sensitività della funzione obiettivo a queste condizioni, come dimostrato nel seguito attraverso differenti casi studio (Capitoli 4 e 5).

3.3 Le caratteristiche del problema di correzione della domanda (caso dinamico)

Il problema di correzione/stima combinata della domanda di trasporto dinamica formulato nel precedente paragrafo, possiede diverse caratteristiche che possono influenzare la scelta della procedura risolutiva da utilizzare.

Queste caratteristiche sono:

1. l'essere un problema a larga scala;
2. la non linearità delle relazioni in esso contenute;
3. la non analiticità;
4. la stocasticità intrinseca ai modelli di assegnazione dinamica.

Per quanto riguarda il primo punto, il problema di stima è un problema a larga scala in quanto le variabili da determinare dipendono dalla dimensione della rete, che nelle applicazioni in contesti reali, può raggiungere dimensioni ragguardevoli. In più, in dinamica, compare l'elemento temporale che va oltremodo ad incrementare la complessità del problema.

Si può generalmente scrivere che:

$$(\mathbf{d}_1^* \dots \mathbf{d}_{n_h}^*) = f(n_{od}, n_h) \quad (3.3)$$

Ovvero, come precedentemente riportato, il numero di variabili da stimare (le singole matrici $\mathbf{d}_1^* \dots \mathbf{d}_{n_h}^*$) sono funzione della dimensione della rete (numero di coppie origine-destinazione) e della suddivisione temporale assunta.

Collegandosi alla variabile “tempo”, entra in gioco un secondo fattore: l'individuazione della dimensione temporale adeguata per lo studio del problema. Di fatti una suddivisione temporale dell'intervallo di analisi eccessivamente fitta produce non solo un aumento della dimensione del problema, ma anche la variabilità e dunque l'instabilità del dato stimato.

La seconda caratteristica del problema di stima è la non linearità delle relazioni in esso contenute: tali relazioni si riferiscono al deflusso veicolare e quindi alle relazioni tra le grandezze flusso-densità e velocità nonché alle relazioni che descrivono il fenomeno della congestione. Inoltre le stesse funzioni di *goodness of fit* presenti nella funzione obiettivo sono solitamente non lineari.

Tale non linearità si traduce in una funzione obiettivo che potenzialmente presenta numerosi minimi locali.

La dipendenza della funzione obiettivo dagli output di un modello di simulazione a larga scala, non lineare, implica che sia impossibile ottenere una formulazione esplicita della funzione stessa e delle sue derivate. La connessione implicita tra i risultati della simulazione ed i parametri incogniti suggerisce un approccio risolutivo che lavori attraverso l'approssimazione di funzioni e che quindi utilizzi il modello di simulazione come una "scatola nera".

L'ultimo fattore interessante riguarda i modelli di assegnazione dinamica che fungono da strumento essenziale nel processo di stima. I modelli di assegnazione, dovendo rappresentare la complessità del mondo reale, sono caratterizzati da diverse fonti di stocasticità. Il problema si pone qualora il simulatore stocastico abbia implicazioni nella formulazione del problema e ciò può accadere quando, ripetendo le simulazioni partendo dagli stessi input, i risultati ottenuti variano. In questo caso non è sufficiente una sola simulazione, ma le simulazioni vanno ripetute al fine di ottenere un valor medio.

3.4 Metodi di ottimizzazione per problemi a larga scala, non lineari e stocastici

I metodi di ottimizzazione per problemi a larga scala, non lineari e stocastici possono essere classificati in (Balakrishna, 2006):

1. *path search*;
2. *pattern search*;
3. *random search*.

Di seguito verranno analizzati singolarmente valutandone pregi e difetti con il fine di identificare il più adatto alla risoluzione del problema di stima della domanda in dinamica intraperiodale.

La notazione generale considerata per un problema di ottimizzazione sarà di seguito:

$$\text{Min}_{\theta} z(\theta) \tag{3.4}$$

soggetto a vincoli di estremo inferiore e superiore:

$$\mathbf{l} \leq \theta \leq \mathbf{u} \tag{3.5}$$

dove θ è un vettore k-dimensionale.

3.4.1 Metodi Path Search

I metodi di tipo *path search* partono da un vettore iniziale di parametri e si muovono in una certa direzione nella speranza di migliorare il valore della funzione obiettivo. Spesso la direzione viene individuata attraverso la valutazione del gradiente. A questi appartengono la Response Surface Methodology (RSM) e l'approssimazione stocastica (Stochastic approximation – SA).

Response Surface Methodology (RSM)

E' un approccio largamente utilizzato basato su un'approssimazione polinomiale della funzione obiettivo in prossimità del valore iniziale dei parametri da stimare. Una volta valutata l'approssimazione, se ne calcola il gradiente e ci si sposta lungo tale direzione. Solitamente viene utilizzata una “superficie approssimante” di tipo lineare o quadratica.

L'applicazione del RSM può essere implementata o in maniera sequenziale o attraverso meta-modelli.

Nel primo caso si costruisce una “superficie approssimante” lineare a ripetizione sino ad ottenere una buona riduzione della funzione obiettivo. Dopodichè si effettua un'approssimazione polinomiale quadratica fino ad annullare il valore del gradiente.

Nel secondo caso si identificano una serie di punti attorno alla soluzione corrente ed i valori della funzione obiettivo per questi punti sono utilizzati per costruire una curva approssimante. Successivamente vengono utilizzati metodi di ottimizzazione deterministica per valutare il gradiente di tale curva e aggiornare la soluzione.

Informazioni dettagliate circa tale metodo possono essere trovate in Kleijnen (1987).

La limitatezza del RSM risiede nell'essersi dimostrato un metodo inefficace nel caso di funzioni obiettivo contenenti elevati picchi o profonde vallate. Un recente metodo RSM che cerca di superare tale limite è SNOBFIT (Huyer e Neumaier, 2004), il quale lavora mantenendo una popolazione di punti e calcolando per ognuno di essi la rispettiva superficie quadratica approssimante. In questo modo l'algoritmo è meno vincolato dalla scelta del punto iniziale, così che sia possibile ricercare la soluzione ottima anche in presenza di situazioni di elevati picchi. Il problema in entrambi i casi risiede nell'elevato numero di valutazioni della funzione obiettivo necessarie per la convergenza e quindi nel conseguente notevole numero di simulazioni richieste.

Approssimazione stocastica (SA)

I metodi basati sull'approssimazione stocastica tracciano una sequenza di soluzioni che vengono di volta in volta aggiornate tramite il calcolo della stima del gradiente \hat{g} . Per la generica iterazione $i+1$:

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i - a^i \hat{\mathbf{g}}(\boldsymbol{\theta}^i) \quad (3.6)$$

A seconda del modo con cui viene calcolata tale stima \hat{g} si può differenziare tra:

- metodi delle differenze finite (Finite Difference – FD);
- Simultaneous Perturbation Stochastic Approximation (SPSA).

Nel caso del metodo FD (Kiefer e Wolfowitz, 1952), due possibili schemi (Figura 3.1:) vengono proposti per il calcolo della stima del gradiente:

1. il *Symmetric Design* secondo cui:

$$\hat{\mathbf{g}}_k(\boldsymbol{\theta}^i) = \frac{z(\boldsymbol{\theta}^i + c^i \mathbf{e}_k) - z(\boldsymbol{\theta}^i - c^i \mathbf{e}_k)}{2c^i} \quad (3.7)$$

2. l'*Asymmetric Design* secondo cui:

$$\hat{\mathbf{g}}_k(\boldsymbol{\theta}^i) = \frac{z(\boldsymbol{\theta}^i + c^i \mathbf{e}_k) - z(\boldsymbol{\theta}^i)}{c^i} \quad (3.8)$$

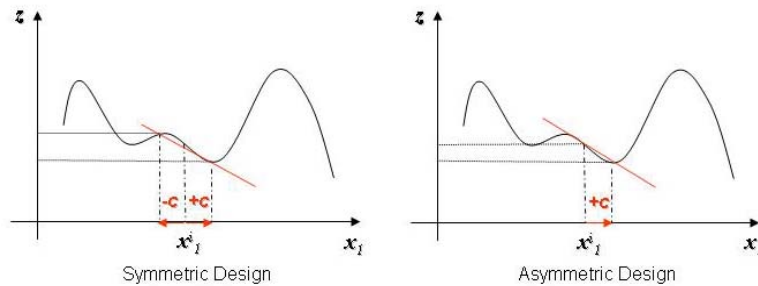


Figura 3.1: *Symmetric e Asymmetric Design*

Tali schemi vanno di volta in volta a calcolare ogni singola componente del gradiente, con il fine di ottenerne il vettore complessivo.

E' possibile notare come nel caso del Symmetric Design siano necessarie $2k$ valutazioni della funzione obiettivo per ogni iterazione per poter

valutare la stima del gradiente; al contrario nell'Asymmetric Design sono sufficienti $k+1$ valutazioni, il che comporta un minore onere computazionale. Questa proprietà diviene particolarmente importante nel caso in cui il numero di iterazioni necessarie alla convergenza sia particolarmente rilevante e qualora i tempi necessari alla valutazione di ogni funzione obiettivo siano eccessivamente lunghi.

Nel SPSA (Spall, 1998, 1999) è possibile risparmiare notevoli risorse computazionali attraverso il calcolo di due sole funzioni obiettivo ad iterazione per ottenere la stima del gradiente. In questo caso infatti la stima non dipende dalle dimensioni del problema ed è:

$$\hat{\mathbf{g}}(\boldsymbol{\theta}^i) = \frac{z(\boldsymbol{\theta}^i + c^i \boldsymbol{\Delta}^i) - z(\boldsymbol{\theta}^i - c^i \boldsymbol{\Delta}^i)}{2c^i} \begin{bmatrix} (\Delta_1^i)^{-1} \\ (\Delta_2^i)^{-1} \\ \vdots \\ (\Delta_k^i)^{-1} \end{bmatrix} \quad (3.9)$$

Dove la distribuzione del vettore $\boldsymbol{\Delta}$ di perturbazioni random di dimensione pari al numero di variabili da stimare (k) deve soddisfare la condizione che i $\{\Delta_k^i\}$ siano indipendenti e simmetricamente distribuiti attorno allo zero con valore finito dell'inverso dei momenti $E(|\Delta_k^i|^{-1})$ per ogni i, k .

Tale fattore porta in letteratura alla scelta della distribuzione simmetrica di Bernoulli (± 1), mentre esclude l'uso della distribuzione uniforme o della normale. Ulteriori condizioni necessarie alla convergenza dell'algoritmo (Spall, 1988, 1992) sono relative agli step c_i , presente nel calcolo della stima del gradiente, ed a_i per l'aggiornamento della soluzione: questi devono convergere verso il valore nullo in modo né troppo rapido, né troppo lento. Infine la funzione z deve essere più volte differenziabile nell'intorno della soluzione $\boldsymbol{\theta}^*$.

Mentre il metodo FD simula l'approccio della direzione di massima discesa, l'SPSA non segue necessariamente una vera direzione di discesa verso l'ottimo. E' comunque dimostrato che i due algoritmi, sotto le condizioni precedentemente riportate, convergono alla stessa soluzione in un numero comparabile di iterazioni (l'errore tra i due diminuisce all'aumentare delle iterazioni, Figura 3.2).

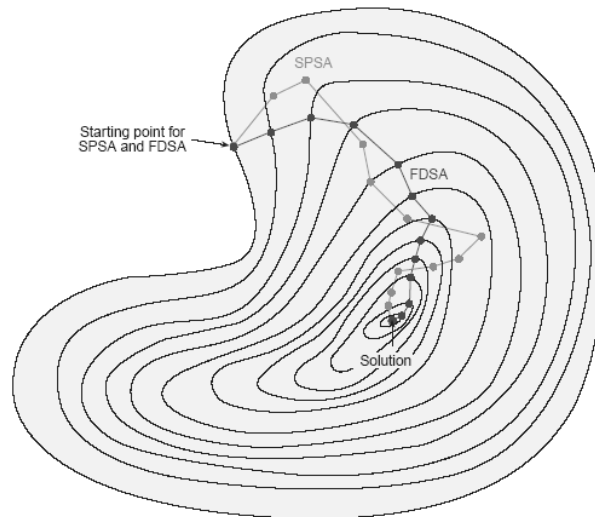


Figura 3.2: Confronto tra la convergenza di SPSA e FDSA

3.4.2 Metodi Pattern Search

I metodi di tipo *pattern search* sono detti metodi di ricerca diretta in quanto non necessitano del calcolo delle derivate e determinano la migliore direzione semplicemente confrontando i valori della funzione obiettivo.

Tra questi:

- Hooke e Jeeves (1961);
- Nelder e Mead (1965);
- Box (1965).

Metodo di Hooke e Jeeves

L'algoritmo di Hooke e Jeeves (1961), partendo da una soluzione iniziale θ^0 , ricerca una "buona direzione" per la prima componente della soluzione stessa. Dunque, mantenendo fisse tutte le altre componenti, si muove di un certo passo in entrambi i lati di θ_1^0 e confronta i valori della funzione obiettivo tra i due punti. Se non viene individuato un miglioramento il passo viene ridotto sino ad individuare una direzione di discesa. Si genera quindi un punto intermedio lungo tale direzione. A partire da questo nuovo punto, il metodo riprende a cercare una "buona direzione" ma per la seconda componente e continua sino ad aver raccolto

tutte le direzioni in un unico vettore. Infine viene aggiornata la soluzione iniziale.

Sebbene la possibilità di poter individuare una direzione di discesa senza valutare il gradiente sembra essere un'alternativa allettante, il metodo riportato si presta ad essere un metodo di ricerca prettamente locale in quanto restringe rapidamente il suo spazio di ricerca in un piccolo intorno della soluzione corrente e non riesce a garantire una soluzione soddisfacente anche nel caso di reti di piccole-medie dimensioni.

Metodo di Nelder e Mead

Il metodo di Nelder e Mead (1965) mantiene una popolazione di $k+1$ possibili soluzioni ed i rispettivi valori delle funzioni obiettivo. Ad ogni iterazione la soluzione peggiore è sostituita dal suo "riflesso" rispetto al centroide calcolato in virtù dei punti rimanenti. La popolazione "collassa" in un unico punto quando viene raggiunta la convergenza. La popolazione iniziale può essere realizzata estraendo in maniera random i punti tra i vincoli inferiore e superiore. In letteratura (Lagarias et al., 1998) è dimostrato come sia difficile verificare una rigorosa convergenza di tale algoritmo quando la dimensione del problema è $k > 2$ e come spesso tale metodo termini nell'individuazione di un punto sub-ottimale in presenza di una funzione obiettivo instabile.

Metodo di Box

Il metodo di Box (1965) è una estensione del Nelder e Mead, ma mentre quest'ultimo mantiene una popolazione di dimensioni $k+1$, nel metodo di Box la popolazione deve essere pari almeno a $2k$. Il metodo si dimostra piuttosto lento nella convergenza e ciò porta a credere di aver individuato la soluzione ottima quando questa non è in realtà ancora stata raggiunta.

3.4.3 Metodi Random Search

I metodi di tipo *random search* sono metodi probabilistici che sfruttano le informazioni derivanti dai passi precedenti per orientare il percorso di ricerca. Tra questi troviamo gli algoritmi evolutivi ed i metodi Simulated Annealing.

Algoritmi Evolutivi

Gli Algoritmi Evolutivi sono dei metodi di ricerca che traggono la loro origine dalle tecniche di selezione naturale esistenti nel mondo biologico.

Essi conducono la ricerca a partire da una “popolazione” di soluzioni. Ad ogni iterazione, prevedono una selezione competitiva che elimina le soluzioni meno valide. Le soluzioni che presentano un maggiore “adattamento” sono “ricombinate” con le altre scambiandosi delle parti di elementi o “mutate” apportando piccoli cambiamenti ad un singolo elemento della soluzione stessa. Gli operatori “ricombinazione” e “mutazione” sono quindi utilizzati per generare nuove soluzioni localizzate nelle regioni dell’insieme di definizione del problema dove sono state già trovati dei punti validi.

Questi algoritmi possono essere classificati in base alla tecnica evolutiva utilizzata e ad essi appartengono gli Algoritmi Genetici particolarmente adatti ai problemi di ottimizzazione di natura combinatoria.

Simulated Annealing

E’ un metodo di ottimizzazione ispirato dall’analogia con il processo fisico di “ricottura” e raffreddamento di un liquido al fine di ottenere la formazione di cristalli cui corrispondono degli stati di minimo dell’energia. La procedura consiste nel raffreddare un liquido inizialmente ad alta temperatura; il raffreddamento avviene lentamente in modo che ad ogni istante il sistema sia approssimativamente in equilibrio termodinamico. Man mano che il processo continua le molecole del liquido diventano sempre più ordinate ed il sistema si avvicina al congelamento cioè allo stato di energia più basso.

Originariamente, la procedura si basava sullo schema *Metropolis* (Metropolis *et al.*, 1953) secondo il quale la configurazione iniziale del sistema, caratterizzata da una data temperatura T ed una data energia E , viene perturbata; il calcolo della variazione di energia dE ad esso relativa permette, se questa è negativa, di accettare il nuovo stato; viceversa, se la variazione di energia è positiva, il nuovo stato verrà accettato con una probabilità data dal fattore di Boltzman $\exp(-dE/T)$. Il processo è ripetuto un numero sufficiente di volte da avere un buon campione statistico per la temperatura T ; successivamente, la temperatura viene diminuita e l’intero processo ripetuto finché non si raggiunge lo stato di congelamento.

Modifiche a questo schema originario hanno portato alla formulazione attuale dell’algoritmo. Esso, ad ogni iterazione, genera delle perturbazioni casuali della attuale soluzione (da θ' a θ''). Queste sono accettate sempre se comportano delle riduzioni del valore della funzione obiettivo (cioè se $z(\theta'') - z(\theta') < 0$); in caso di incremento, la nuova soluzione viene accettata con una probabilità P data da:

$$\exp(-(z(\boldsymbol{\theta}^{\prime\prime}) - z(\boldsymbol{\theta}^{\prime}))/T) \quad (3.10)$$

In altre parole, la nuova soluzione è accettata se la probabilità P è maggiore o uguale della probabilità R rappresentata da un numero estratto casualmente nell'intervallo $[0,1)$; se R è minore di P , la nuova configurazione è accettata come punto di partenza per il passo successivo. E' la probabilità di muoversi in salita che permette all'algoritmo di non restare intrappolato in una "vallata" di minimo locale. Inoltre, ripetendo questa procedura molte volte, le configurazioni accettate assumono la forma di una distribuzione di Boltzman alla temperatura considerata.

3.5 Procedura risolutiva

Il problema di correzione della matrice O/D dinamica viene di seguito risolto attraverso l'applicazione del metodo di ottimizzazione di tipo "path search" SPSA (Simultaneous Perturbation Stochastic Approximation, Spall 1998): questo lavora tramite un'approssimazione del gradiente e può trovare una buona soluzione quando il punto di partenza (domanda target) è sufficientemente vicino alla soluzione ottima. Nella presente ricerca si ipotizza per l'appunto di partire da una domanda target sufficientemente buona, ipotesi che deriva, come riportato nel paragrafo 3.1, da un presupposto ingegneristico che vede il continuo aggiornamento della suddetta informazione per la comune pratica professionale. Il passo successivo, scopo del presente lavoro, è quello di poter, a partire dalla domanda target, stimare la domanda all'orizzonte temporale attuale, considerando l'evoluzione del sistema delle attività in un arco temporale ristretto (non si tratta dunque di previsioni a lungo termine). In tale ottica l'applicazione di un metodo che sostanzialmente si basa sulla nozione di gradiente può essere considerato corretto.

Inoltre l'SPSA permette un notevole risparmio dal punto di vista computazionale, fattore indispensabile per orientarsi verso una stima simultanea della domanda in contesti reali anche nel caso off-line.

Partendo dalla generica iterazione k , la domanda all'iterazione $k+1$ può essere calcolata come:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - a_k \bar{\mathbf{g}}(\mathbf{x}_k) \quad (3.11)$$

dove:

a_k = step progressivo di aggiornamento della matrice all'iterazione k
dell'algoritmo

$\bar{\mathbf{g}}(\mathbf{x}_k)$ = gradiente approssimato medio all'iterazione k

Il gradiente approssimato medio $\bar{\mathbf{g}}(\mathbf{x}_k)$ è calcolato come la media di m stime del gradiente:

$$\bar{\mathbf{g}}(\mathbf{x}_k) = \text{media}_m \{ \hat{\mathbf{g}}_m(\mathbf{x}_k) \} \quad (3.12)$$

Ogni stima $\hat{\mathbf{g}}_m(\mathbf{x}_k)$ è basata sulla perturbazione simultanea di ogni componente di \mathbf{x}_k , così che tutti gli elementi della matrice siano perturbati insieme per ottenere due nuove matrici attraverso le quali valutare la funzione obiettivo. Nel caso in cui la perturbazione avvenga in maniera simmetrica (Symmetric Design – SD), ovvero la matrice \mathbf{x}_k sia perturbata simmetricamente in eccesso ed in difetto, si ha:

$$\hat{\mathbf{g}}_m(\mathbf{x}_k) = \frac{z((\mathbf{x}_1 \dots \mathbf{x}_{n_h})_k + c_k \Delta_m) - z((\mathbf{x}_1 \dots \mathbf{x}_{n_h})_k - c_k \Delta_m)}{2c_k} \begin{bmatrix} (\Delta_m^1)^{-1} \\ (\Delta_m^2)^{-1} \\ \vdots \\ (\Delta_m^{n_v})^{-1} \end{bmatrix} \quad (3.13)$$

Dove la distribuzione del vettore Δ di perturbazioni random di dimensione n_v (con $n_v = n_h \times n_{od}$) deve soddisfare la condizione che i $\{\Delta_m^i\}$ siano indipendenti e simmetricamente distribuiti attorno allo zero con valore finito dell'inverso dei momenti $E(|\Delta_m^i|^{-1})$ per ogni m, i .

La procedura risolutiva si articola nei seguenti punti (Figura 3.3):

1. per la generica iterazione k , si definiscono i valori degli step a_k e c_k e si estrae il vettore delle variabili random Δ_m ;
2. partendo dalla matrice \mathbf{x}_k , si definiscono due ulteriori matrici a partire dal passo c_k e dal vettore Δ_m , una matrice “superiore” \mathbf{x}_k^+ ed una matrice “inferiore” \mathbf{x}_k^- (SD);
3. si calcola il valore della funzione obiettivo per le due matrici \mathbf{x}_k^+ ed \mathbf{x}_k^- , così come definita nella (3.1);

4. si calcola un'approssimazione del gradiente $\hat{\mathbf{g}}_m(\mathbf{x}_k)$ e si ripete tale calcolo per un numero m di volte al fine di valutare il gradiente approssimato medio $\bar{\mathbf{g}}(\mathbf{x}_k)$
5. si aggiorna la matrice di domanda secondo la (3.11);
6. si verifica un prefissato criterio di convergenza e se la verifica ha avuto esito positivo, l'algoritmo termina, altrimenti si ritorna al punto 1.

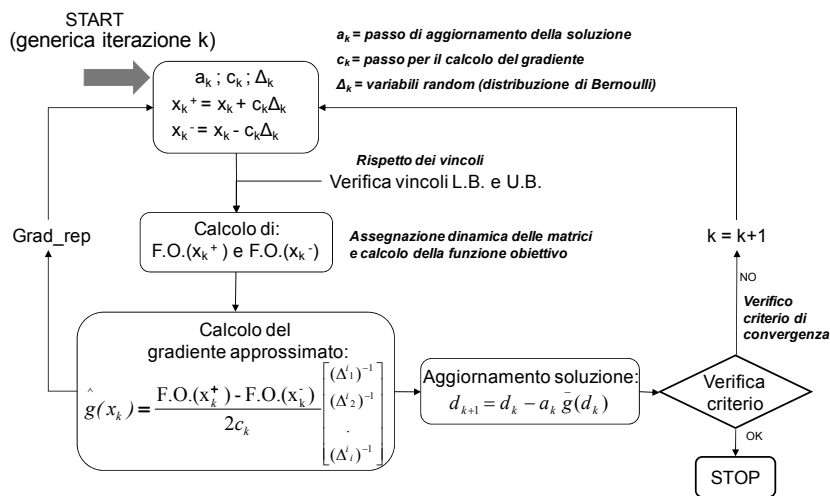


Figura 3.3: Algoritmo SPSA applicato alla correzione della matrice OD

La formulazione base dell'algoritmo SPSA adotta una regola predefinita per l'aggiornamento dello step di ricerca lineare lungo la direzione di discesa (Spall, 1998), ovvero:

$$a_k = \frac{a}{(A + k)^\alpha} \quad (3.14)$$

dove:

k = iterazione dell'algoritmo;

a = valore iniziale del passo di aggiornamento della domanda;

A, α = coefficienti non negativi.

3.5.1 Aggiornamento della soluzione corrente (Spall, GSM, PI)

L'algoritmo SPSA, così come illustrato nel paragrafo 3.5, è stato testato su di una rete test di piccole dimensioni (Figura 3.4), considerando una domanda complessiva di partenza pari a 2.500 veic/h (domanda target), distribuita in un unico intervallo temporale e semplificando la funzione obiettivo con il fine di considerare i soli conteggi di traffico.

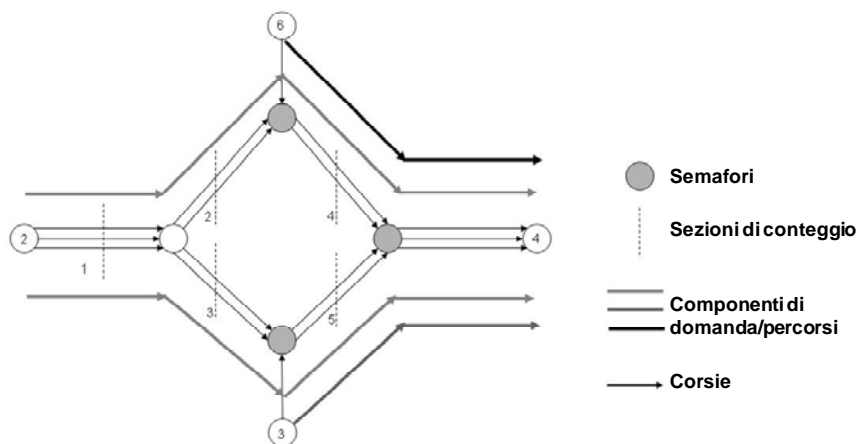


Figura 3.4: Rete test

Cinque sezioni di conteggio sono state posizionate sulla rete; la presenza di tre semafori con stesso ciclo ($C=120$ sec) e verde ugualmente distribuito tra i diversi approcci ha permesso di incrementare la congestione.

Nella situazione analizzata, seppur particolarmente semplificata, si è potuto notare che l'utilizzo della formulazione riportata in (3.14), con tutta la complessità derivante dalla scelta dei coefficienti A ed α , porta ad una convergenza irregolare della funzione obiettivo con risultati non soddisfacenti persino dopo numerose iterazioni (Figura 3.5).

Questa scarsa efficacia ha portato a considerare ulteriori metodi di ricerca lineare della soluzione come il metodo della sezione aurea (Golden Section Method, GSM), il quale ha dimostrato, sotto le stesse ipotesi di offerta e domanda analizzate in precedenza, la propria abilità nell'ottenere una buona convergenza in un ridotto numero di iterazioni (Figura 3.6, Figura 3.7).

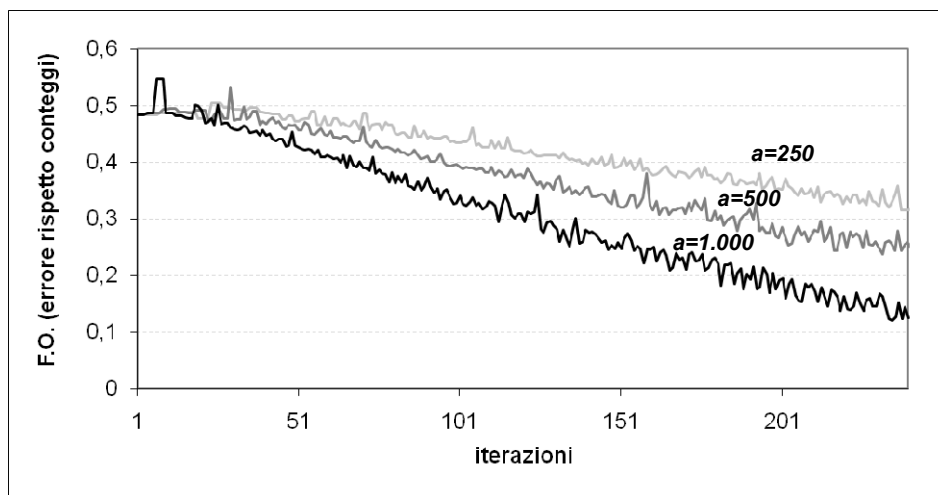


Figura 3.5: Andamento della funzione obiettivo per rete test in condizioni di domanda e offerta semplificate utilizzando lo step classico

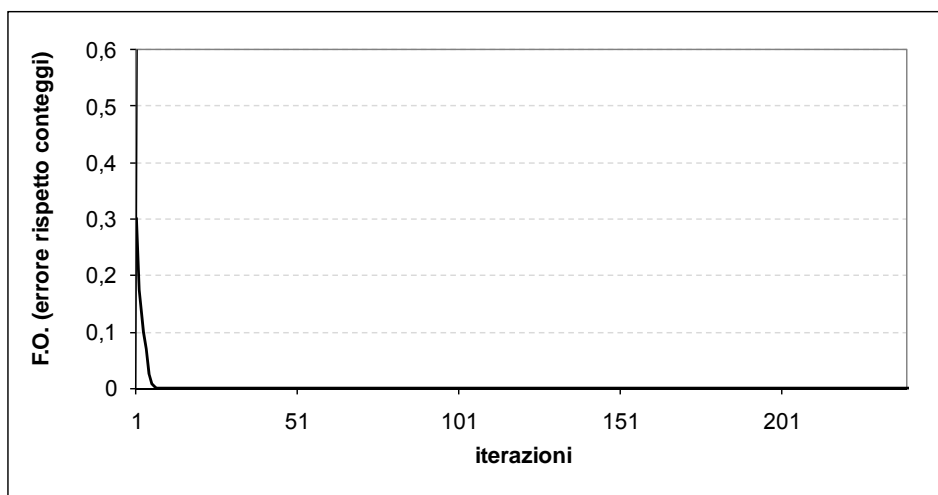


Figura 3.6: Applicazione del metodo della sezione aurea per rete test in condizioni di domanda e offerta semplificate

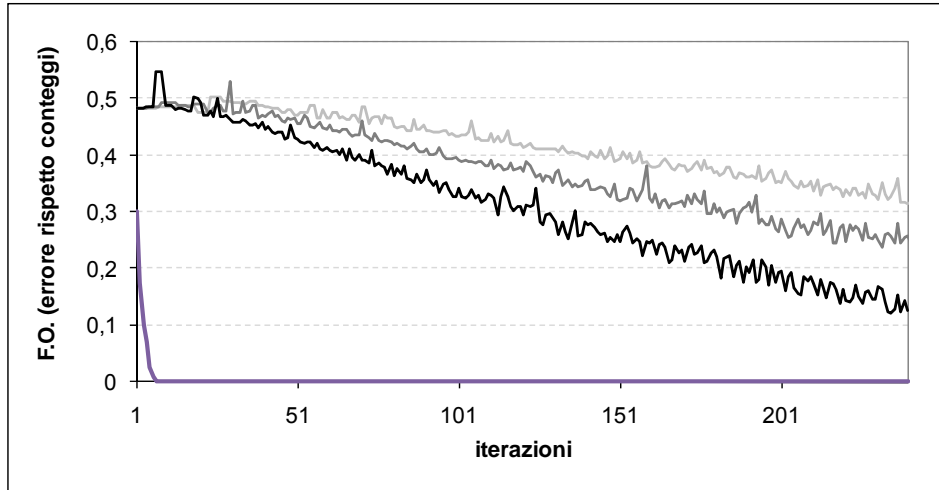


Figura 3.7: Confronto funzione obiettivo con utilizzo step classico e GSM

Ovviamente il miglioramento e la stabilità ottenuti nella convergenza dell'algoritmo di correzione della matrice attraverso l'uso del metodo della sezione aurea producono un incremento dei tempi computazionali necessari ad ogni iterazione.

Difatti il metodo GSM esplora un intervallo chiuso al fine di ricercarne il valore ottimo della funzione obiettivo. E' una procedura iterativa basata su una riduzione sequenziale dell'intervallo di incertezza iniziale e richiede 4 punti di partenza (la domanda alla generica iterazione k ed ulteriori tre matrici individuate lungo la direzione di discesa), il relativo valore della funzione obiettivo ed un nuovo punto (ed una funzione obiettivo) ad ogni iterazione (Figura 3.8).

Per ridurre i tempi computazionali, mantenendo comunque una procedura che converga alla soluzione cercata in tempi ragionevoli, nel presente studio si è analizzata l'eventuale sostituzione del metodo della sezione aurea con un'approssimazione polinomiale (Polynomial Interpolation – PI). A partire dai quattro punti, ovvero dalle quattro matrici, così come definite per il metodo della sezione aurea, si calcola l'espressione del polinomio di terzo grado interpolante. Da tale funzione è possibile ricavare analiticamente il relativo punto di minimo, che verrà considerato come soluzione sub-ottimale ad ogni iterazione (Figura 3.9).

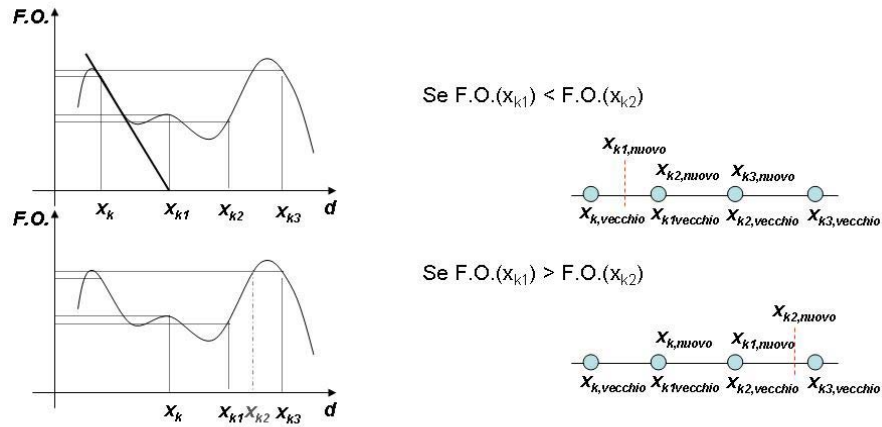


Figura 3.8: Il metodo della sezione aurea (GSM)

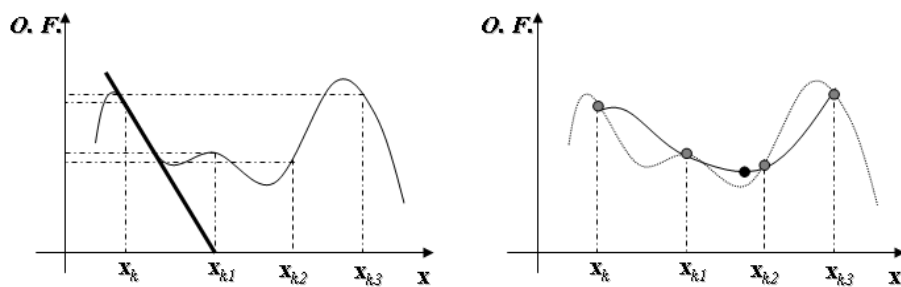


Figura 3.9: Applicazione dell'interpolazione polinomiale per il calcolo della soluzione lungo la direzione di discesa

L'individuazione degli ulteriori tre punti a partire dalla soluzione corrente, lungo la direzione di discesa, determina l'intervallo in cui ricercare il punto di minimo; in particolare le estremità di tale intervallo sono regolate dalla soluzione corrente e dal terzo punto, la cui scelta determina la maggiore o minore dimensione dell'area di ricerca e dunque la maggiore o minore "dimensione della ricerca locale".

3.5.2 Gradiente approssimato (SD – AD)

Nel presente studio, sono state introdotte ulteriori modifiche all'algoritmo SPSA con il fine di ridurre il numero di valutazioni della funzione obiettivo (e quindi il numero di assegnazioni dinamiche - Dynamic Traffic Assignments o DTAs) richieste ad ogni iterazione.

In particolare si è proposta la sostituzione del metodo Symmetric design (SD), normalmente utilizzato per il calcolo del gradiente, con il metodo Asymmetric design (AD) il quale prevede la perturbazione del punto iniziale in una sola direzione e quindi la valutazione di una sola funzione obiettivo.

Asymmetric design:

$$\hat{\mathbf{g}}_m(\mathbf{x}_1 \dots \mathbf{x}_{n_h})_k = \frac{z(\mathbf{x}_1 \dots \mathbf{x}_{n_h})_k + c_k \Delta_m - z(\mathbf{x}_1 \dots \mathbf{x}_{n_h})_k}{c_k} \begin{bmatrix} (\Delta_m^1)^{-1} \\ (\Delta_m^2)^{-1} \\ \cdot \\ (\Delta_m^{n_h})^{-1} \end{bmatrix} \quad (3.15)$$

In questo modo si vanno a dimezzare il numero di valutazioni della funzione obiettivo richieste in ogni perturbazione simultanea delle componenti.

3.6 Strumenti di ricerca (Dynameq)

Le simulazioni dinamiche di traffico richieste per calcolare la funzione obiettivo sono state effettuate tramite l'utilizzo del software DYNAMEQ (Florian *et al.* 2006).

DYNAMEQ è un modello di assegnazione dinamica (Dynamic Traffic Assignment - DTA) che utilizza alcune varianti del metodo del gradiente ed il metodo delle medie successive (Method of Successive Averages - MSA) per determinare le scelte di percorso in una condizione di equilibrio dinamico.

Le scelte di percorso sono modellate come variabili decisionali governate dal principio di "ottimo dell'utente" secondo il quale ogni utente della rete punta a minimizzare i suoi costi e quindi il tempo di viaggio sul percorso utilizzato. Tutti gli utenti hanno accesso alle informazioni della rete, ovvero ai tempi di viaggio su tutti i percorsi (utilizzati e non utilizzati).

L'algoritmo risolutivo prende la forma di una procedura iterativa realizzata per convergere verso queste condizioni ed è caratterizzata da due componenti principali:

1. un metodo per determinare un nuovo set di flussi di percorso variabili nel tempo, data l'esperienza maturata circa i tempi di percorrenza nell'iterazione precedente;
2. un metodo per determinare i flussi di arco ed i tempi di percorrenza risultanti da un dato set di flussi di percorso in ingresso. Quest'ultimo si riferisce alla fase di caricamento della rete, che è svolta attraverso un efficiente modello di simulazione del traffico ad eventi. Il modello rappresenta esplicitamente i sistemi di regolazione presenti e cattura realisticamente fenomeni di congestione, quali la formazione delle code ed il loro propagarsi da arco in arco (spill-back).

3.6.1 Rappresentazione della rete, sistemi di controllo e domanda

La rete di trasporto è definita da un insieme di archi e nodi. Ad ogni arco sono associate una serie di misure: lunghezza, numero di corsie e velocità libera.

Ad ogni nodo, ogni manovra di svolta viene definita univocamente tramite il proprio arco uscente ed entrante: sono note il numero di corsie utilizzate per la manovra, nonché la velocità massima e la capacità per corsia.

Il modello non richiede informazioni dettagliate sulla geometria di ogni intersezione, quali la sua dimensione e forma o il raggio di curvatura delle manovre. Ogni corsia di un arco ed ogni manovra possono essere ristrette ad un sottoinsieme di classi veicolari, permettendo la modellazione della corsie di tipo HOV (high-occupancy vehicle), o corsie riservate per autobus o taxi.

Il modello permette inoltre la specificazione di informazioni dettagliate sul controllo semaforico, andando così a rappresentare semafori predeterminati e ramp metering.

La domanda è definita per ogni classe veicolare attraverso una serie di matrici OD variabili nel tempo. Ogni classe veicolare comprende una o più tipologie di veicolo, che si distinguono per attributi fisici quali la lunghezza effettiva del veicolo ed il tempo di risposta del guidatore.

Questi parametri permettono insieme di definire la densità critica e la velocità dell'onda associate ad ogni tipologia di veicolo. Inoltre, per una data velocità libera dell'arco, questi attributi forniscono la capacità di saturazione associata a quel veicolo su quell'arco (o manovra).

3.6.2 Il modello di assegnazione dinamica

Due differenti approcci vengono generalmente utilizzati per simulare il comportamento di scelta dell'utente: un'assegnazione dinamica "en route" ed un'assegnazione dinamica all'equilibrio. In DYNAMEQ l'approccio utilizzato è quello di cercare una soluzione approssimata alle condizioni di equilibrio.

Nel problema di assegnazione all'equilibrio vengono considerate solo le scelte di percorso di tipo "pre-trip", ovvero le scelte effettuate dall'utente prima di aver intrapreso il suo percorso. Le scelte di percorso sono modellate come variabili decisionali governate dal principio di "ottimo dell'utente" secondo il quale ogni utente della rete punta a minimizzare i suoi costi e quindi il tempo di viaggio sul percorso utilizzato. Tutti gli utenti hanno accesso alle informazioni della rete, ovvero ai tempi di viaggio su tutti i percorsi (utilizzati e non utilizzati). L'algoritmo risolutivo prende la forma di una procedura iterativa realizzata per convergere verso queste condizioni.

L'approccio risolutivo adottato per risolvere il modello dinamico di equilibrio (3.16)-(3.18) è basato su una discretizzazione temporale

$\tau = 1, 2, \dots, \left\lfloor \frac{T_d}{\Delta t} \right\rfloor$, in cui Δt è la durata scelta dell'intervallo temporale. Il

risultato è un modello discreto nel tempo.

L'enunciazione matematica del problema di equilibrio dinamico discretizzato nel tempo trova luogo nello spazio dei flussi di percorso h_k^τ per tutti i percorsi k appartenenti all'insieme K_i per ogni coppia origine-destinazione $i \in I$, per ogni τ . I flussi di percorso nella regione ammissibile Ω soddisfano i vincoli di non negatività e di conservazione del flusso:

$$\Omega^\tau = \left\{ h_k^\tau : \sum_{k \in K_i} h_k^\tau = g_i^\tau, i \in I, \forall \tau; \quad h_k^\tau \geq 0, k \in K_i, i \in I, \forall \tau \right\} \quad (3.16)$$

Ed una versione temporale del principio di Wardrop risulta nel modello:

$$\begin{aligned}
h_k^\tau &\in \Omega, \quad u_i^\tau = \min_{k \in K_i} \{s_k^\tau(t)\} \\
s_k^\tau(t) &= u_i^\tau \quad \text{se } h_k^\tau > 0 \\
s_k^\tau(t) &\geq u_i^\tau \quad \text{se } h_k^\tau = 0 \quad \text{per tutti } i \in I, k \in K_i, \tau = 1, 2, \dots, \left\lfloor \frac{T_d}{\Delta t} \right\rfloor
\end{aligned} \tag{3.17}$$

che risulta equivalente al risolvere la disuguaglianza variazionale discretizzata:

$$\sum_{\tau} \sum_{k \in K} s_k^\tau(h_k^\tau)(h_k^\tau - h_k^{\tau^*}) \geq 0, \quad K = \bigcup_{i \in I} K_i \tag{3.18}$$

La dimostrazione dell'esistenza e dell'unicità della soluzione di questo modello dipende dalle proprietà della mappa $s(h/g)$, ovvero della dipendenza tra tempi di percorrenza d'arco e di percorso e tra flussi di percorso e domanda. Poiché le proprietà di questa mappa non sono semplici da verificare, in quanto tali relazioni sono l'output del modello di simulazione e non di una funzione analitica, nessuna affermazione può essere fatta circa l'esistenza e l'unicità della soluzione.

Il principio di equilibrio è quindi utilizzato come guida nel calcolare una soluzione approssimata della disuguaglianza variazionale discretizzata.

3.6.3 L'algoritmo risolutivo

L'algoritmo risolutivo utilizzato in DYNAMIQ consiste di due componenti principali oltre che della fase di calcolo dei cammini minimi dinamici:

1. un metodo per determinare un nuovo set di flussi di percorso variabili nel tempo, data l'esperienza maturata circa i tempi di percorrenza nell'iterazione precedente;
2. un metodo per determinare i flussi di arco ed i tempi di percorrenza risultanti da un dato set di flussi di percorso in ingresso.

L'algoritmo richiede inoltre un set iniziale di flussi di percorso. La struttura generale dell'algoritmo è riportata in Figura 3.10.

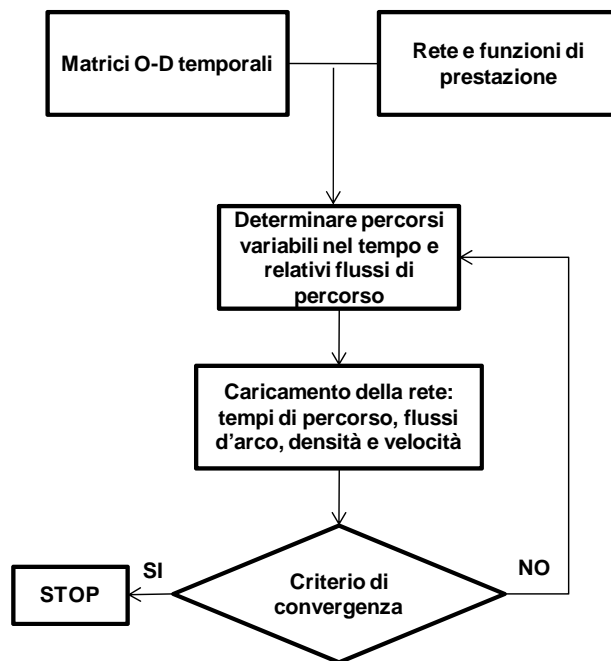


Figura 3.10: Struttura dell'algorithmo risolutivo del modello di assegnazione dinamica

I flussi di percorso di input h_k^r , $k \in K$, sono determinati attraverso una variante del metodo delle medie successive (MSA), applicato ad ogni coppia O-D ed intervallo temporale τ .

Un insieme iniziale di soluzioni possibili è calcolato assegnando la domanda per ogni intervallo temporale su di un set di successivi cammini minimi. Dalla seconda iterazione e sino ad un predefinito numero massimo N di iterazioni, i tempi di percorrenza d'arco variabili nel tempo dopo ogni caricamento sono utilizzati per determinare un nuovo set di cammini minimi dinamici che vengono aggiunti al set di cammini correnti.

All'iterazione n , $n \leq N$, il volume assegnato su ogni percorso appartenente al set iniziale è posto pari a $\frac{g_i^r}{n}$. Dopodiché per ogni

iterazione m , $m > N$, sono identificati i cammini minimi tra quelli effettivamente usati ed i flussi di percorso vengono ridistribuiti tra i percorsi conosciuti.

Se il flusso su di un particolare percorso scende al di sotto di un valore predeterminato, il percorso non viene considerato ed il suo flusso viene distribuito sugli altri percorsi utilizzati. Questo approccio euristico è simile all'algoritmo di Lawgphonpanich e Hearn (1984) per la soluzione del modello di equilibrio statico a domanda fissa.

L'algoritmo viene riassunto di seguito (Figura 3.11):

- *Step 0 Inizializzazione ($l=1$):*
Calcola i cammini minimi per i diversi intervalli temporali utilizzando i tempi di percorrenza a flusso nullo; carica la domanda per ottenere una soluzione iniziale; $l=l+1$
- *Step 1 Riallocazione dei flussi sui percorsi:*
Step 1.1 Se $l \leq N$
calcola un nuovo cammino minimo dinamico;
assegna su ogni percorso k il flusso $\frac{g_i^\tau}{l}$.
- Step 1.2 Se $l > N$*
identifica i percorsi più brevi tra quelli utilizzati;
ridistribuisce i flussi come segue:

$$h_k^l(\tau) = \begin{cases} h_k^{l-1}(\tau) \left(\frac{l-1}{l} \right) + \frac{g_i^\tau}{l}, & \text{se } s_k^l(\tau) = u_k^l(\tau) \\ h_k^{l-1}(\tau) \left(\frac{l-1}{l} \right) & \text{altrimenti} \end{cases}$$

- *Step 2 Criterio di arresto:*
Se $l \geq l_{max}$ o $RGap \leq \varepsilon \rightarrow \text{STOP}$
Altrimenti ritorna allo Step 1

Nonostante non possa esistere alcuna prova formale sulla convergenza di questo algoritmo, dato che la mappa di carico della rete non è espressa in maniera analitica, per quantificare la bontà della soluzione raggiunta viene utilizzata una misura di confidenza ispirata dai modelli di equilibrio statico. Questa misura è la differenza tra il tempo complessivo di viaggio maturato ed il tempo complessivo che si avrebbe se tutti i veicoli avessero scelto il cammino minimo corrente.

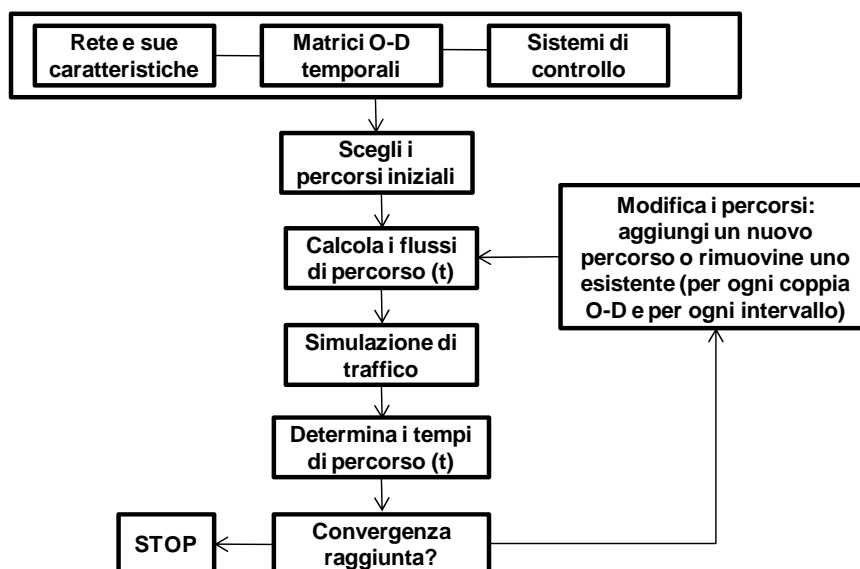


Figura 3.11: Algoritmo di assegnazione dinamica all'equilibrio in Dynameq

Perciò la misura di confidenza (Relative Gap) per ogni intervallo di partenza τ può essere calcolata come:

$$RGap^{\tau}(n) = \frac{\sum_{i \in I} \sum_{k \in K_i} h_k^{\tau}(n) s_k^{\tau}(n) - \sum_{i \in I} g_i^{\tau} u_i^{\tau}(n)}{\sum_{i \in I} g_i^{\tau} u_i^{\tau}(n)} \quad (3.19)$$

dove $u_i^{\tau}(n)$ sono le lunghezze dei cammini minimi all'iterazione n . Un "Relative Gap" pari a zero indica un perfetto equilibrio dinamico. Chiaramente questo è un obiettivo difficile da raggiungere in qualsiasi assegnazione dinamica.

E' molto importante notare come questo modello, anche se nella sua formulazione generale si avvicina molto ai modelli macroscopici di deflusso, è in effetti un modello di tipo discreto. La procedura di caricamento della rete, realizzata attraverso una simulazione ad eventi, muove i singoli veicoli sugli archi della rete.

3.6.4 Il modello di caricamento della rete

Il cuore del modello di simulazione è il seguente modello semplificato di veicolo accodato:

$$x_f(t) = \min[x_f(t - \varepsilon) + \varepsilon V, x_l(t - R) - L] \quad (3.20)$$

in cui $x_f(t)$ è la traiettoria di un veicolo (funzione del tempo), L è la lunghezza effettiva del veicolo, R è il tempo di reazione dell'utente, V è la velocità libera ed ε è un intervallo temporale arbitrariamente piccolo. I pedici f ed l denotano rispettivamente le traiettorie del veicolo che segue e che precede (follower – leader). Questo modello considera solo la velocità libera (senza vincoli di accelerazione) combinata con un semplice modello di impedimento della collisione e può facilmente essere rappresentato attraverso il ben noto diagramma triangolare flusso-densità, anche noto come diagramma fondamentale del traffico (Mahut, 2000).

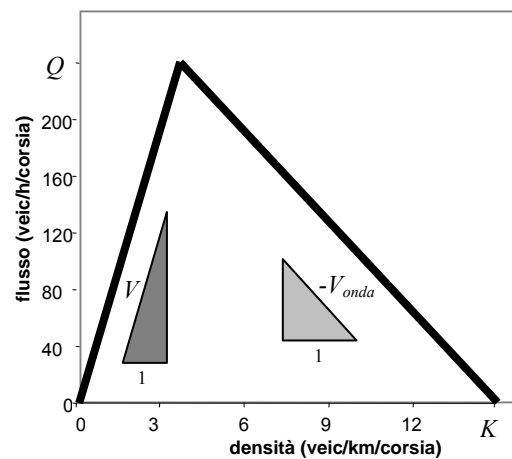


Figura 3.12: Diagramma triangolare flusso-densità

Questa relazione può essere risolta in maniera rigorosa, così come è possibile calcolare il tempo in cui un veicolo entra ed esce da ogni arco utilizzando la seguente formula:

$$t_n(0) = \max \left[t_n(-X_1) + \frac{X_1}{V_1}, t_{n-1}(0) + \left(R + \frac{L}{V_2} \right), t_{n-X_2/L}(X_2) + R \left(\frac{X_2}{L} \right) \right] \quad (3.21)$$

Dove X_1 ed X_2 sono le lunghezze degli archi di monte e di valle rispetto alla posizione $x=0$, V_1 e V_2 sono le velocità libere rispettivamente su questi due archi ed i pedici indicano i veicoli numerati in ordine sequenziale. Questa “formulazione d’arco” fornisce un modo molto pratico e computazionalmente efficiente per modellare il traffico senza calcolare analiticamente le variabili di stato (posizione, velocità, etc.) di ogni veicolo in ogni istante. Si può notare come questa espressione possa essere applicata restrittivamente al caso di un arco con un’unica corsia.

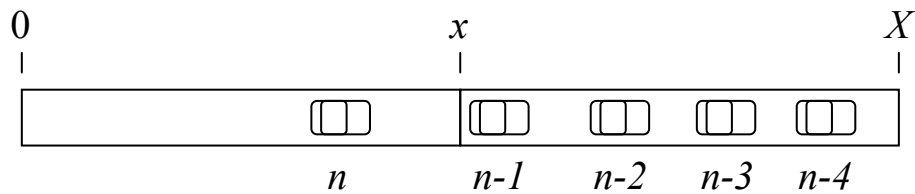


Figura 3.13: Modello semplificato di veicolo accodato

Nel caso di più corsie, è necessario riferirsi ad una diversa formulazione della (3.21) che può calcolare il tempo di ingresso e di uscita di ogni veicolo (Figura 3.13), ma anche catturare l’interazione tra veicoli dovuta al cambiamento di corsia (Mahut 2000). Il modello per più corsie sviluppa una serie di procedure euristiche per modellare la scelta della corsia da parte dell’utente, che tiene conto del percorso desiderato dall’utente a valle della sua posizione corrente, le corsie che possono essere utilizzate per compiere il movimento successivo di svolta e le condizioni prevalenti di traffico su ogni corsia tra il guidatore e la fine dell’arco.

Il modello d’arco suddetto determina l’istante in cui il veicolo desidera entrare nell’arco come funzione dello storico dei tempi di uscita e di ingresso. L’informazione è utilizzata in modo da utilizzare rigorosamente il precedente modello di veicolo accodato.

In sintesi, il modello di veicolo accodato è esteso in maniera ricorsiva per essere applicato ad una sequenza di veicoli. Per esempio, piuttosto che modellare la relazione tra il veicolo 1 ed il veicolo 2 (il numero più alto segue il numero più basso) e la seguente relazione tra il veicolo 2 ed il veicolo 3, il modello permette di esprimere direttamente l’interazione tra il veicolo 1 ed il veicolo 3. Ciò può essere esteso a qualsiasi numero di veicoli, che è l’essenza dell’espressione d’arco riportata precedentemente.

Da un punto di vista concettuale, siccome la prima causa del ritardo in una rete è dovuta alle intersezioni, il ruolo della dinamica in tale modello è quello di propagare correttamente i ritardi intercorsi tra i veicoli a monte ed a valle, ovvero laddove questi ritardi possono influenzare i tempi di ingresso dei veicoli nell'arco. Anziché modellare esplicitamente la posizione di ogni veicolo in modo da determinare quando la congestione dell'arco inizia ad influenzare questi tempi di ingresso, i ritardi all'uscita sono propagati direttamente in ingresso. Questa particolare caratteristica – l'abilità di risolvere rigorosamente il modello di traffico sull'intero arco – è stata dimostrata anche nel caso di modello di onda cinematica basata sul diagramma fondamentale triangolare (Newell, 1993).

Questo modello dinamico del traffico – che può essere caratterizzato come un modello continuo nel tempo, continuo nello spazio, con flusso discretizzato – è combinato con un modello di nodo che rappresenta esplicitamente i sistemi semaforici predeterminati e modella anche le interazioni tra veicoli alle intersezioni non semaforizzate attraverso una logica di *gap-acceptance*. Il sistema combinato è poi risolto utilizzando un algoritmo a eventi che permette di modellare l'intera rete. I modelli ad eventi sono fondamentalmente differenti dai modelli per intervalli temporali discreti, nel senso che:

- i modelli ad intervalli temporali discreti prevedono che ad ogni step temporale (generalmente 1 secondo o meno), tutti i dati richiesti per ogni veicolo in ogni singolo modello (veicolo accodato, cambiamento di corsia, *gap-acceptance*, scelta del percorso) siano aggiornati e gli output (accelerazione, decelerazione etc.) siano ricalcolati;
- i modelli ad evento prevedono che i singoli sottomodelli siano aggiornati solo qualora uno degli input cambi in maniera rilevante. Dunque i singoli sottomodelli sono tali da garantire che i loro output rimangano validi fino a che i dati di input si mantengono costanti. Un evento viene generato in uno specifico punto qualora venga a variare l'informazione. Generalmente il compimento di un evento risulta nella creazione di uno o più sequenze di eventi.

Il risultato in questo caso è un modello ad eventi molto efficiente da un punto di vista computazionale, ma che rispetta anche le leggi base del traffico e rappresenta esplicitamente il meccanismo di congestione che si presenta nella realtà.

Tutto ciò risulta in un maggior risparmio dei tempi di calcolo rispetto ai modelli di simulazione ad intervalli temporali (di circa uno o due ordini di grandezza).

I tempi computazionali sono di particolare importanza nel contesto dei modelli di assegnazione dinamica, i quali devono effettuare numerose iterazioni per risolvere l'assegnazione d'equilibrio. Il fatto che i modelli di assegnazione dinamica debbano essere applicati su reti di grandi dimensioni rende il fattore temporale ancora più importante.

Di seguito è riportata una sintesi delle proprietà del modello di simulazione analizzato ed utilizzato per la ricerca:

1. vengono modellate le manovre dei singoli veicoli, offrendo una rappresentazione realistica del meccanismo di congestione (soprattutto ai nodi) e della sua propagazione laterale (tra le corsie) e longitudinale (lungo gli archi);
2. è utilizzata una regola di selezione della corsia che utilizza la logica del look-ahead;
3. un algoritmo di prevenzione del grid-lock identifica gli archi molto prossimi ad incorrere in tale fenomeno ed interviene sui flussi in ingresso nei cicli dove si presenta questo rischio come fosse un sistema semaforico di tipo adattivo.
4. i tempi computazionali sono molto contenuti se paragonati ad un modello ad intervalli temporali discreti.

4. Applicazione rete test

Il metodo SPSA è stato applicato sulla rete test presentata nel paragrafo 3.5.1. (Figura 3.4) costituita da 8 archi ed 8 nodi, considerando una domanda target composta da tre coppie O-D per un complessivo di 4.000 veic/h, un periodo di simulazione di 50 minuti e differenti ipotesi sulla distribuzione temporale della domanda:

- 2 intervalli temporali da 25 minuti l'uno;
- 5 intervalli temporali da 10 minuti l'uno.

Cinque sezioni di conteggio sono state inserite nella rete; tre intersezioni sono semaforizzate con stesso ciclo ($C=120s$) e tempo di verde distribuito egualmente tra gli approcci al fine di aumentare la congestione.

Rispetto alla formulazione classica del SPSA, sono state introdotte alcune varianti sia nella fase di calcolo della stima del gradiente che nell'individuazione della nuova soluzione lungo la direzione di discesa. Tutte le combinazioni testate in relazione alle ipotesi effettuate sulla domanda e sulla procedura risolutiva sono riportate nella Tabella 4.1.

Tabella 4.1: *Differenti ipotesi testate del SPSA sulla rete test*

Numero di intervalli temporali	Calcolo del gradient/aggiornamento della soluzione			
	<i>SD/GSM</i>	<i>SD/PI</i>	<i>AD/GSM</i>	<i>AD/PI</i>
2	<i>SD/GSM</i>	<i>SD/PI</i>	<i>AD/GSM</i>	<i>AD/PI</i>
5	<i>SD/GSM</i>	<i>SD/PI</i>	<i>AD/GSM</i>	<i>AD/PI</i>
<i>SD=Symmetric Design</i> <i>AD=Asymmetric design</i> <i>GSM=Golden Section Method</i> <i>PI=Polynomial Interpolation</i>				

La domanda “reale”, valore al quale ci si aspetta che la procedura di correzione converga, è nota in tale ambito. Infatti i presenti esperimenti sono test di laboratorio in cui è possibile ipotizzare un valore di domanda “reale” che, assegnata alla rete tramite gli strumenti di simulazione (Dynameq), fornisca i valori osservati di flusso e velocità in corrispondenza delle sezioni di conteggio. La domanda “reale” viene quindi utilizzata solo inizialmente per generare le misure di traffico, per poi scomparire all'interno del processo di correzione in cui entra in gioco solo la domanda target. Al termine della procedura è possibile riutilizzare

la domanda “reale” per validare i risultati ottenuti. Ovviamente questa validazione è possibile solo in tali esperimenti di laboratorio, in quanto, per un’applicazione standard, la matrice di domanda “reale” non può essere nota e le sole informazioni su di essa sono le informazioni desunte dalle misure di traffico raccolte sulla rete.

Per la valutazione del processo di correzione in termini di corrispondenza rispetto alla domanda “reale” ed alle misure d’arco (flussi e velocità) è stata utilizzata, oltre al valore del R^2 , la statistica RME, così definita:

$$RME(\text{relative mean error}) = \frac{\sum_{ij} |x_{ij}^* - x_{ij}|}{\sum_{ij} x_{ij}} \quad (4.1)$$

dove

x^* : valori delle variabili stimate (domanda stimata dalla procedura e corrispondenti valori di flussi e velocità simulate alle sezioni di conteggio);

x : valori “reali” delle variabili (domanda “reale” e corrispondenti valori di flussi e velocità misurate alle sezioni di conteggio);

i : intervallo temporale;

j : codice identificativo della variabile (coppia O-D o sezione di conteggio).

Iniziando dalla dimensione temporale minore (2 intervalli temporali), i risultati mostrano la validità della metodologia proposta. La funzione obiettivo si stabilizza dopo 25-30 iterazioni, ad eccezione della variante AD-PI, per la quale le iterazioni richieste per raggiungere la stabilità aumentano a circa 90 (Figura 4.1).

Comunque una dettagliata analisi dei valori dei singoli termini presenti nella funzione obiettivo mostra che se la differenza percentuale tra conteggi di traffico e flussi simulati e tra velocità misurate e velocità simulate viene considerata come criterio di convergenza (ovvero qualora l’errore percentuale su ogni arco sia inferiore ad un prefissato valore ϵ), la convergenza dell’algoritmo necessita di un maggior numero di iterazioni quando viene utilizzato il metodo della sezione aurea (GSM, Tabella 4.2).

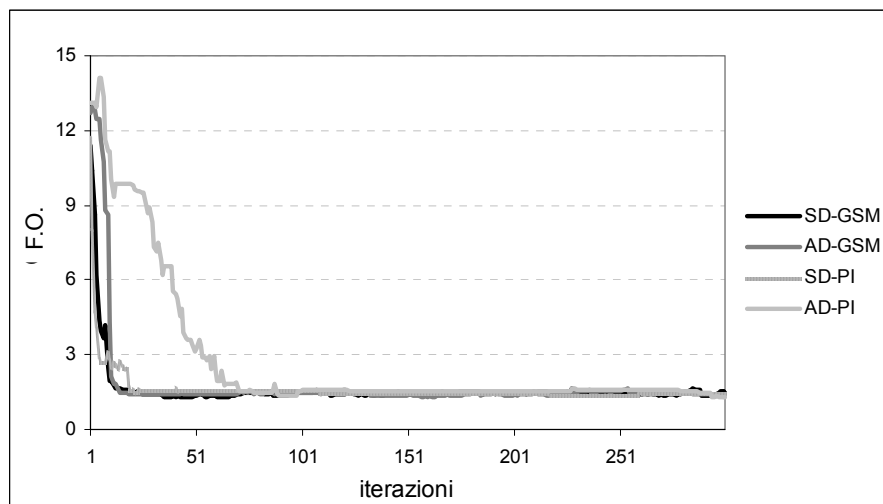


Figura 4.1: Andamento della funzione obiettivo (2 intervalli temporali)

Tabella 4.2: Confronto tra le procedure (2 intervalli temporali)

	<i>Iterazioni per la stabilizzazione della F.O.</i>	<i>Iterazioni per la convergenza</i>
<i>SPSA SD-GSM</i>	25	126
<i>SPSA AD-GSM</i>	25	179
<i>SPSA SD-PI</i>	30	42
<i>SPSA AD-PI</i>	90	99

I risultati a convergenza (Tabella 4.3, Tabella 4.4) mostrano una corrispondenza ottima sia in termini di flussi (di domanda e d'arco) che di velocità: i miglioramenti in termini della statistica RME superano in tutti i casi l'80% rispetto al valore iniziale, mentre il R^2 è sempre prossimo all'unità.

La maggiore efficienza viene raggiunta dal metodo SPSA SD-PI, dove attraverso 336 assegnazioni dinamiche si giunge al risultato illustrato.

Tabella 4.3: Valori iniziali degli indicatori R^2 ed RME

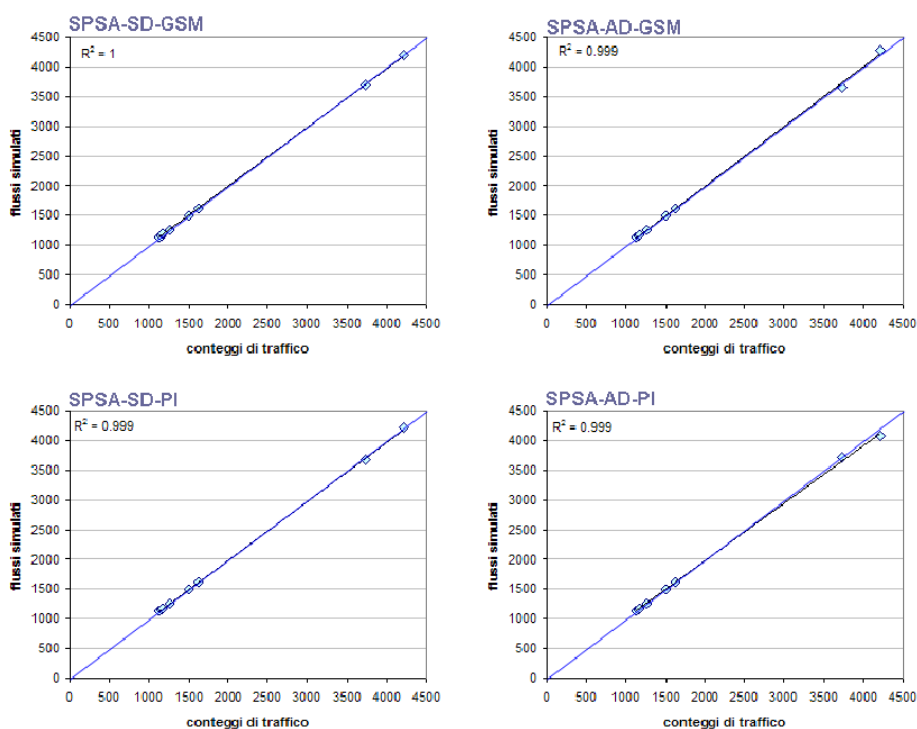
	<i>2 intervalli temporali</i>		
	<i>Domanda</i>	<i>Flussi</i>	<i>Velocità</i>
<i>RME</i>	0,349	0,201	0,516
R^2	0,899	0,787	0,752

Tabella 4.4. Risultati a convergenza (2 intervalli temporali)

	Convergenza			DTA*
	Variazione RME / valore finale R^2			
	Domanda	Flussi	Velocità	
SPSA SD-GSM	-83%/0,996	-98%/1	-91%/0,990	1.512
SPSA AD-GSM	-82%/0,995	-95%/0,999	-92%/0,992	1.791
SPSA SD-PI	-88%/0,997	-98%/0,999	-92%/0,992	336
SPSA AD-PI	-90%/0,998	-96%/0,999	-91%/0,995	595

* Numero di assegnazioni dinamiche

Tutti i risultati precedentemente esposti sono riportati graficamente in termini di corrispondenza rispetto ai valori “reali” nelle figure seguenti (Figura 4.2 per la corrispondenza rispetto ai flussi d’arco misurati – conteggi di traffico, Figura 4.3 per la corrispondenza rispetto alle velocità misurate, Figura 4.4 per la corrispondenza rispetto alla domanda “reale”).

**Figura 4.2:** Corrispondenze conteggi - flussi simulati (2 intervalli temporali)

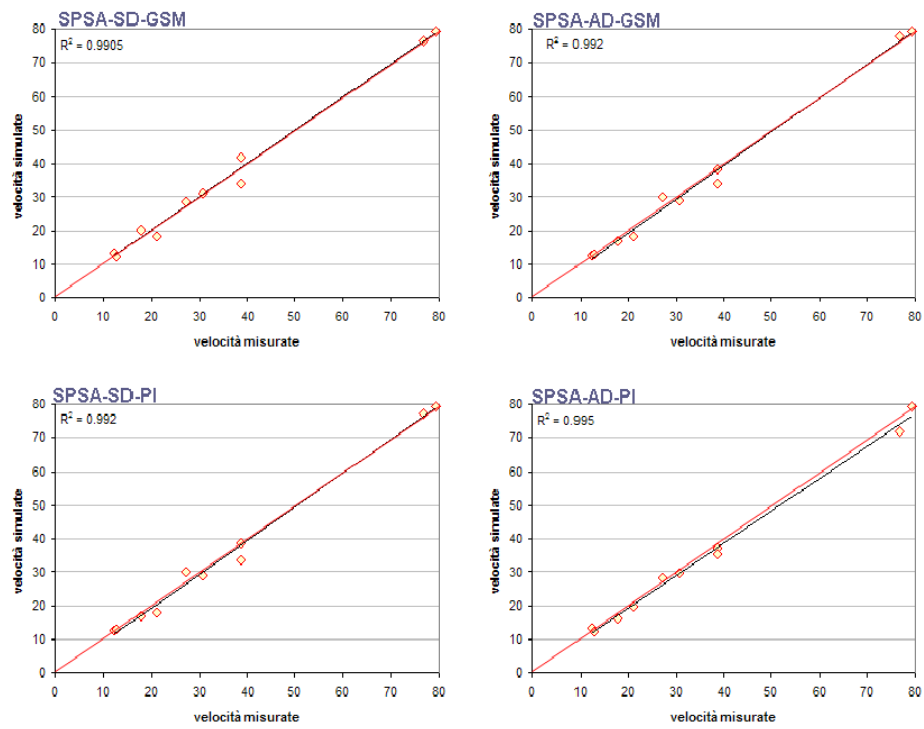
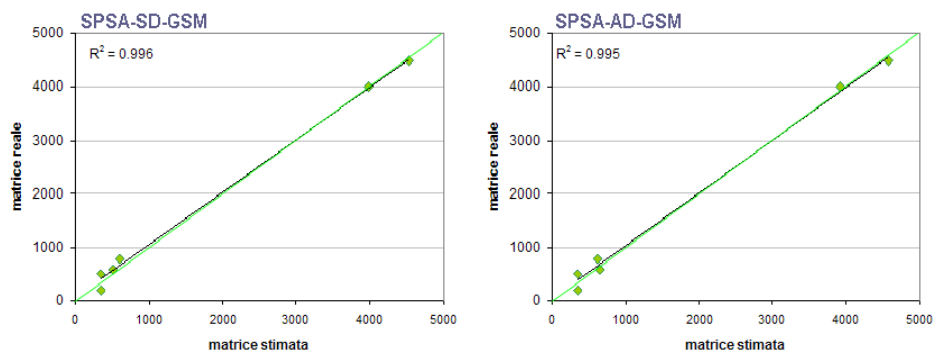


Figura 4.3: Corrispondenze velocità misurate-simulate (2 intervalli temporali)



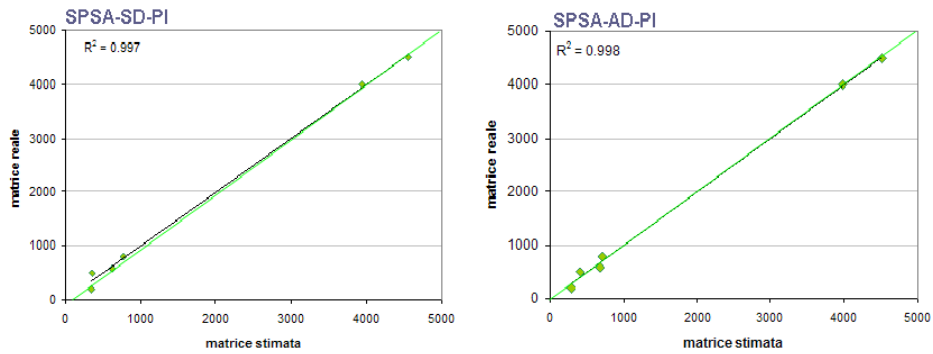


Figura 4.4: Corrispondenze domanda stimata-reale (2 intervalli temporali)

Nel caso di 5 intervalli temporali, la stabilizzazione della funzione obiettivo sembra richiedere un maggior numero di iterazioni qualora si adotti la variante dell'Asymmetric Design per la valutazione del gradiente (AD-GSM/AD-PI). In particolare nel caso AD-GSM, la funzione obiettivo rimane "intrappolata" in un range di valori pari a 10-15 e solo dopo un alto numero di iterazioni (180) salta rapidamente verso una buona soluzione (Figura 4.5).

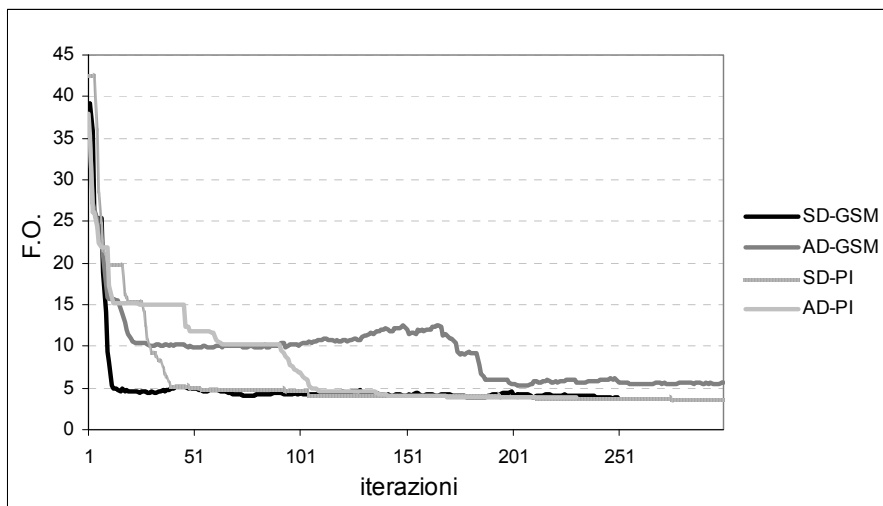


Figura 4.5: Andamento della funzione obiettivo (5 intervalli temporali)

Guardando invece il numero di iterazioni necessarie alla convergenza, si osserva che non c'è più differenza tra l'uso di AD o di SD: infatti in tutti i casi sono necessarie circa 200 iterazioni (Tabella 4.5).

Tabella 4.5: Confronto tra le procedure (5 intervalli temporali)

	Iterazioni per la stabilizzazione della F.O.	Iterazioni per la convergenza
<i>SPSA SD-GSM</i>	25	233
<i>SPSA AD-GSM</i>	190	207
<i>SPSA SD-PI</i>	60	276
<i>SPSA AD-PI</i>	110	202

In termini di corrispondenza su flussi e velocità, sono stati ottenuti buoni risultati in tutti i casi. Valori migliori del R^2 si hanno utilizzando l'*SPSA SD-PI*, ma quest'ultimo richiede un maggior numero di iterazioni (Tabella 4.6, Tabella 4.7).

Tabella 4.6: Valori iniziali degli indicatori R^2 ed RME

	5 intervalli temporali		
	Domanda	Flussi	Velocità
RME	0,349	0,202	0,463
R^2	0,915	0,801	0,668

Tabella 4.7: Risultati a convergenza (5 intervalli temporali)

	Convergenza			
	Variazione RME / valore finale R^2			DTA*
	Domanda	Flussi	Velocità	
<i>SPSA SD-GSM</i>	-46%/0,941	-72%/ 0,948	-87%/0,974	4.194
<i>SPSA AD-GSM</i>	-40%/0,921	-68%/ 0,941	-82%/0,986	2.692
<i>SPSA SD-PI</i>	-58%/0,963	-76%/ 0,971	-84%/0,976	4.968
<i>SPSA AD-PI</i>	-53%/0,918	-74%/ 0,950	-89%/0,967	1.819

* Numero di Assegnazioni Dinamiche

Tutti i risultati precedentemente esposti sono riportati graficamente in termini di corrispondenza rispetto ai valori "reali" nelle figure seguenti (Figura 4.6 per la corrispondenza rispetto ai flussi d'arco misurati –

conteggi di traffico, Figura 4.7 per la corrispondenza rispetto alle velocità misurate, Figura 4.8 per la corrispondenza rispetto alla domanda “reale”).

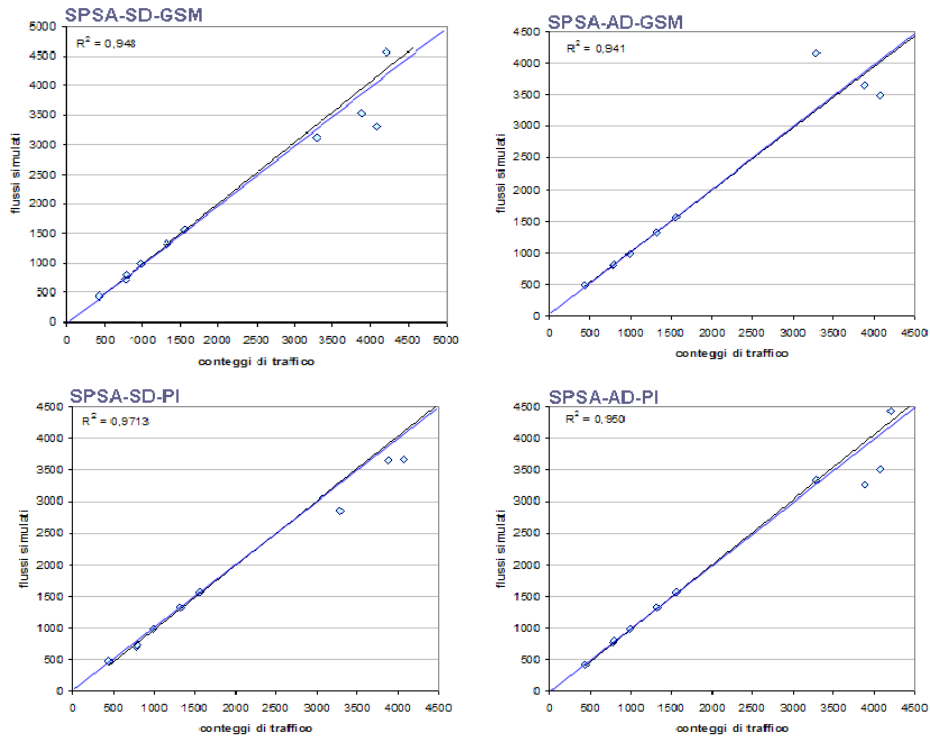
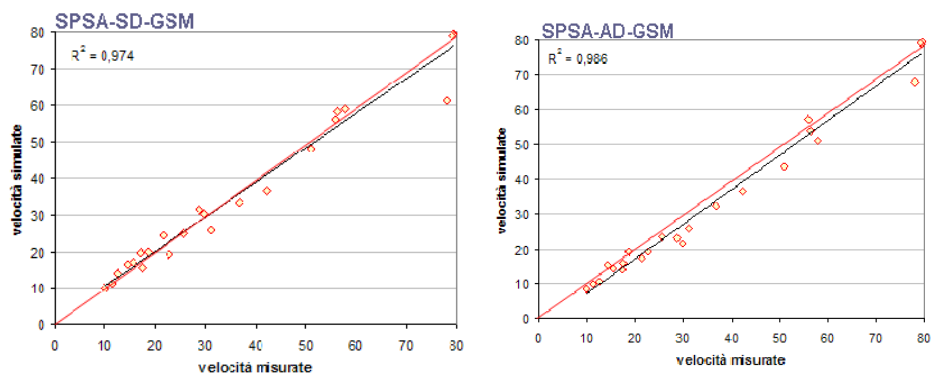


Figura 4.6: Corrispondenze conteggi - flussi simulati (5 intervalli temporali)



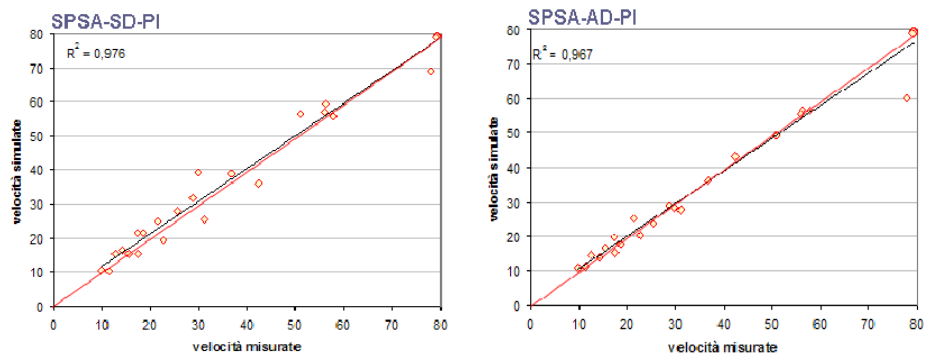


Figura 4.7: Corrispondenze velocità misurate-simulate (5 intervalli temporali)

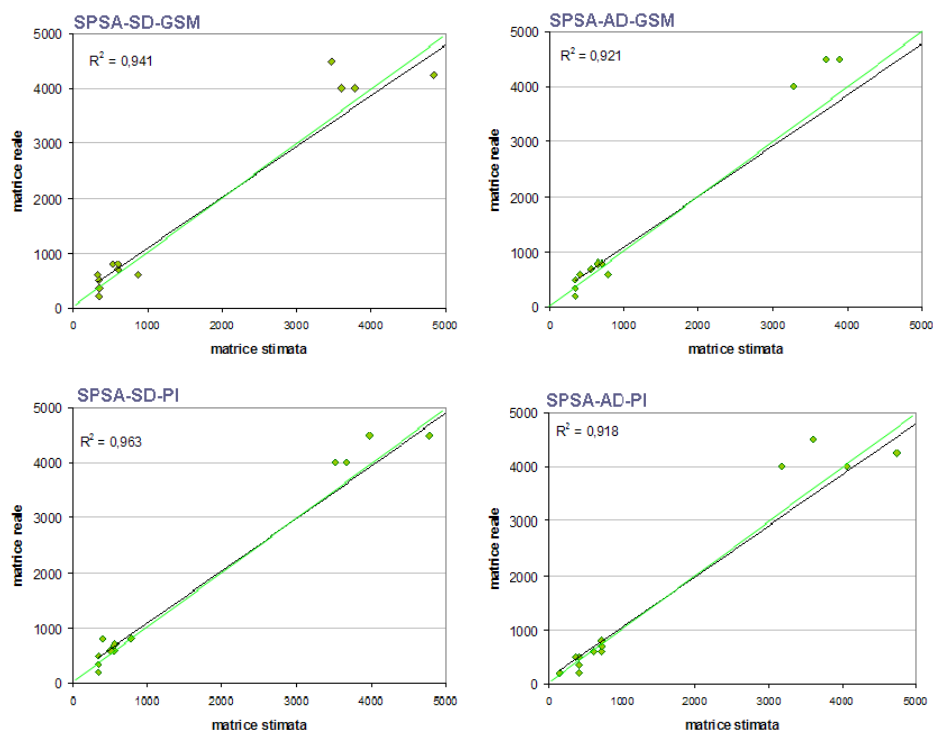


Figura 4.8: Corrispondenze domanda stimata-reale (5 intervalli temporali)

L'SPSA AD-PI ha dimostrato in entrambi i casi (2 e 5 intervalli temporali) la sua abilità nel trovare una buona soluzione. In più, nel caso

di 5 intervalli temporali e quindi quando la dimensione del problema va ad aumentare, esso richiede un minor numero di assegnazioni. Così, considerando la necessità di applicare la procedura proposta per risolvere il problema di correzione della matrice O-D dinamica in particolare in un contesto reale, l'SPSA AD-PI sembra dimostrarsi un buon compromesso tra efficienza ed efficacia.

4.1 Analisi di sensitività

Nel seguito è stata condotta una dettagliata analisi delle potenzialità del metodo SPSA AD-PI nel caso della rete test con 5 intervalli temporali, utilizzando differenti valori della domanda target (da una domanda complessiva di 2.000 veic/h a 9.000 veic/h) come input della procedura di correzione.

Per analizzare i risultati ottenuti sono state utilizzate differenti statistiche, oltre alle già citate R^2 ed RME, applicate sia sui valori della domanda che sui valori di flusso e velocità.

$$MSE \text{ (mean square error)} = \frac{\sum_{ij} (x_{ij}^* - x_{ij})^2}{n_{ij}} \quad (4.2)$$

$$RRMSE \text{ (relative root mean square error)} = \frac{\sqrt{n_{ij} \sum_{ij} (x_{ij}^* - x_{ij})^2}}{\sum_{ij} x_{ij}} \quad (4.3)$$

dove:

x^* : valori stimati delle variabili (domanda stimata dalla procedura e corrispondenti valori di flussi e velocità simulate alle sezioni di conteggio);

x : valori "reali" delle variabili (domanda "reale" e corrispondenti valori di flussi e velocità misurate alle sezioni di conteggio);

n_{ij} : numero delle variabili;

i : intervallo temporale;

j : codice identificativo della variabile (riferito alla coppia O-D o alla sezione di conteggio).

I casi studio affrontati possono essere raggruppati in funzione del diverso livello di congestione che si sviluppa sulla rete a seguito

dell'assegnazione della domanda target rispetto alle condizioni "reali" (ovvero rispetto al livello di congestione che si avrebbe assegnando la domanda "reale"), ed in particolare:

1. stesso livello di congestione delle condizioni "reali" (Test 1);
2. livello di congestione inferiore alle condizioni "reali" (Test da 2 a 7);
3. maggiore livello di congestione che nelle condizioni "reali" (Test 8).

Nel primo caso (Test 1, Tabella 4.8 - Tabella 4.9) i risultati mostrano che l'SPSA AD-PI può portare a stimare una buona domanda se la target e la "reale" sono tra loro abbastanza vicine in termini di congestione sulla rete (buona corrispondenza tra flussi e velocità anche nel punto iniziale della procedura di correzione): ciò dimostra, come ci si poteva aspettare, l'ottima validità del metodo utilizzato come metodo di ottimizzazione locale.

Tabella 4.8: Risultati dell'analisi di sensitività (R^2 ed MSE)

		R^2			MSE		
		Corrispondenza rispetto a:			Corrispondenza rispetto a:		
		Dom. reale	conteggi	velocità	Dom. reale	conteggi	velocità
1	Rif.	0,915	0,801	0,668	$84 \cdot 10^4$	$46 \cdot 10^4$	531
	Ris.	0,936	0,950	0,977	$21 \cdot 10^4$	$70 \cdot 10^3$	15
	Var.	2,30%	18,63%	46,23%	74,5%	84,81%	97,17%
2	Rif.	0,739	0,610	0,456	$17 \cdot 10^5$	$83 \cdot 10^4$	615
	Ris.	0,579	0,737	0,858	$20 \cdot 10^5$	$37 \cdot 10^4$	94
	Var.	-21,7%	20,7%	88,2%	-18%	55%	84,7%
3	Rif.	0,588	0,591	0,774	$19 \cdot 10^5$	$71 \cdot 10^4$	396
	Ris.	0,641	0,747	0,956	$14 \cdot 10^5$	$35 \cdot 10^4$	28
	Var.	9,03%	26,44%	23,42%	23,5%	50,73%	92,94%
4	Rif.	0,464	0,270	0,744	$25 \cdot 10^5$	$12 \cdot 10^5$	315
	Ris.	0,537	0,451	0,960	$20 \cdot 10^5$	$80 \cdot 10^4$	24
	Var.	15,81%	66,84%	29,11%	21,2%	35,63%	92,44%
5	Rif.	0,599	0,303	0,358	$27 \cdot 10^5$	$18 \cdot 10^5$	1097
	Ris.	0,664	0,729	0,939	$12 \cdot 10^5$	$35 \cdot 10^4$	38
	Var.	10,96%	140,9%	162,5%	53,7%	80,58%	96,51%
6	Rif.	0,515	0,352	0,534	$24 \cdot 10^5$	$14 \cdot 10^5$	656
	Ris.	0,464	0,354	0,595	$25 \cdot 10^5$	$13 \cdot 10^5$	511
	Var.	-9,91%	0,47%	11,45%	-3,8%	1,49%	22,08%
7	Rif.	0,754	0,610	0,611	$15 \cdot 10^5$	$91 \cdot 10^4$	836
	Ris.	0,678	0,636	0,791	$16 \cdot 10^5$	$81 \cdot 10^4$	273
	Var.	-10,1%	4,19%	29,27%	-5,8%	10,82%	67,32%
8	Rif.	0,948	0,629	0,382	$31 \cdot 10^5$	$62 \cdot 10^4$	559
	Ris.	0,926	0,922	0,998	$25 \cdot 10^4$	$99 \cdot 10^3$	2
	Var.	-2,4%	46,7%	161,2%	92,2%	83,9%	99,6%

Tabella 4.9: Risultati dell'analisi di sensitività (RME ed RRMSE)

		RME			RRMSE		
		Corrispondenza rispetto a:			Corrispondenza rispetto a:		
		Dom. reale	conteggi	velocità	Dom. reale	conteggi	velocità
1	Rif.	0,349	0,202	0,463	0,519	0,362	0,593
	Ris.	0,163	0,053	0,049	0,262	0,141	0,100
	Var.	53,17%	73,67%	89,44%	49,5%	61,02%	83,19%
2	Rif.	0,478	0,269	0,440	0,742	0,485	0,638
	Ris.	0,504	0,124	0,153	0,806	0,325	0,250
	Var.	-5,40%	54,01%	65,13%	-8,7%	32,92%	60,82%
3	Rif.	0,424	0,229	0,404	0,772	0,448	0,512
	Ris.	0,338	0,096	0,057	0,675	0,315	0,136
	Var.	20,37%	57,84%	85,84%	12,5%	29,81%	73,43%
4	Rif.	0,513	0,293	0,321	0,895	0,593	0,457
	Ris.	0,437	0,167	0,099	0,795	0,476	0,126
	Var.	14,88%	43,04%	69,25%	11,2%	19,77%	72,51%
5	Rif.	0,602	0,514	0,676	0,927	0,717	0,852
	Ris.	0,460	0,117	0,127	0,631	0,316	0,159
	Var.	23,57%	77,20%	81,28%	31,9%	55,93%	81,33%
6	Rif.	0,538	0,392	0,497	0,882	0,620	0,659
	Ris.	0,528	0,384	0,424	0,899	0,615	0,582
	Var.	1,85%	2,05%	14,80%	-1,9%	0,75%	11,73%
7	Rif.	0,435	0,324	0,589	0,694	0,507	0,744
	Ris.	0,449	0,288	0,318	0,714	0,478	0,425
	Var.	-3,34%	11,30%	46,02%	-2,9%	5,57%	42,83%
8	Rif.	0,649	0,184	0,346	1,003	0,419	0,608
	Ris.	0,174	0,070	0,025	0,281	0,168	0,038
	Var.	73,2%	62,2%	92,7%	72%	59,9%	93,7%

Nel secondo caso (Test da 2 a 7, Tabella 4.8 - Tabella 4.9) l'algoritmo migliora sempre la corrispondenza rispetto ai conteggi ed alle velocità misurate e, specialmente in termini di velocità d'arco, riesce a lavorare in maniera particolarmente efficace raggiungendo valori del R^2 superiori allo 0,8; anche i miglioramenti relativi ai flussi d'arco possono essere molto forti, ma in alcuni casi non soddisfacenti ai fini del raggiungimento della domanda "reale" (è questo il caso dei test 3, 4 e 5, Tabella 4.8 - Tabella 4.9, dove la domanda stimata al termine della procedura è sempre migliore della domanda target, ma non ancora confrontabile con il valore "reale"). Inoltre in alcuni casi è possibile ottenere un peggioramento della corrispondenza rispetto alla domanda "reale" anche se i risultati della simulazione riflettono molto bene le misure di traffico (Test 2 e 7, Tabella 4.8 - Tabella 4.9).

Solo in un caso (Test 6, Tabella 4.8 - Tabella 4.9) l'algoritmo non raggiunge i risultati sperati per ogni componente analizzata (peggiore della domanda stimata e miglioramenti poco rilevanti di flussi e velocità).

I risultati precedentemente illustrati in relazione al caso in cui la domanda target riporta un minore livello di congestione della domanda “reale”, possono essere spiegati attraverso le seguenti osservazioni:

- esiste una sostanziale differenza dal punto di vista numerico tra il termine legato ai flussi ed il termine legato alle velocità all’interno della funzione obiettivo; tale differenza è dovuta al fatto che i flussi possono variare su di un intervallo più vasto di valori (maggior limite superiore), mentre le velocità variano tra il valore nullo e la velocità libera. La conseguenza di questo fattore è che l’algoritmo si trova a lavorare in maniera differente sui due termini;
- la distanza rispetto alla domanda target, fondamentale nella formulazione della funzione obiettivo per limitare il problema della indeterminatezza della procedura di stima, vincola la ricerca in una determinata regione dello spazio delle soluzioni. In questo caso la soluzione finale dipende sostanzialmente dal “livello di bontà” associato dalla domanda target;
- una buona corrispondenza in termini di misure di traffico (flussi e velocità) non assicura una domanda stimata vicina alla domanda “reale”; infatti nella formulazione stessa del problema di correzione non esiste altra informazione sulle condizioni reali (ed in particolare sulla domanda) se non tramite le misure di traffico ed uno stesso set di misure su di un sottoinsieme di archi della rete può derivare da matrici di domanda anche molto diverse tra loro.

Per superare le suddette limitazioni, qualora la matrice target non sia sufficientemente vicina alla reale, ogni termine della funzione obiettivo deve fornire un adeguato contributo al processo di correzione: ciò implica che principalmente gli errori sugli archi, che rappresentano le distanze tra flussi/velocità osservati/e e simulati/e, devono essere pesati in maniera appropriata (normalizzazione attraverso l’uso di pesi esterni):

$$(\mathbf{d}_1^* \dots \mathbf{d}_{n_h}^*) = \arg \min_{(\mathbf{x}_1 \dots \mathbf{x}_{n_h}) \geq 0} \left[f_1(\mathbf{x}_1 \dots \mathbf{x}_{n_h}, \hat{\mathbf{d}}_1 \dots \hat{\mathbf{d}}_{n_h}) + \beta \cdot (f_2(\mathbf{v}_1 \dots \mathbf{v}_{n_h}, \hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_{n_h})) + \alpha \cdot f_3(\mathbf{s}_1 \dots \mathbf{s}_{n_h}, \hat{\mathbf{s}}_1 \dots \hat{\mathbf{s}}_{n_h}) \right] \quad (4.4)$$

I pesi (α, β) possono essere determinati alla prima iterazione dell’algoritmo, ovvero quando è possibile stimare i valori dei differenti termini della funzione obiettivo.

Nel caso di matrice target con minor congestione rispetto alle condizioni di traffico reali, i test condotti normalizzando i termini della funzione obiettivo con il fine di bilanciare i rispettivi pesi, hanno dimostrato ottimi risultati rispetto al medesimo caso senza pesi (8-9% di miglioramento della statistica R^2 rispetto alla domanda, 50-100% su flussi e velocità, Tabella 4.10, Tabella 4.12).

Tabella 4.10: Funzione Obiettivo con pesi (R^2 e MSE)

		R^2			MSE		
		Corrispondenza rispetto a:			Corrispondenza rispetto a:		
		Dom. reale	conteggi	velocità	Dom. reale	conteggi	velocità
2	Rif.	0,739	0,610	0,456	$17 \cdot 10^3$	$83 \cdot 10^4$	615
	Ris.	0,579	0,737	0,858	$20 \cdot 10^3$	$37 \cdot 10^4$	94
	Var.	-21,7%	20,7%	88,2%	-18%	55%	84,7%
2 con pesi	Rif.	0,739	0,610	0,456	$17 \cdot 10^3$	$83 \cdot 10^4$	615
	Ris.	0,802	0,858	0,933	$75 \cdot 10^4$	$19 \cdot 10^4$	42
	Var.	8,44%	40,65%	104,6%	56,2%	76,7%	93,26%
6	Rif.	0,515	0,352	0,534	$24 \cdot 10^3$	$14 \cdot 10^3$	656
	Ris.	0,464	0,354	0,595	$25 \cdot 10^3$	$13 \cdot 10^3$	511
	Var.	-9,91%	0,47%	11,45%	-3,8%	1,49%	22,08%
6 con pesi	Rif.	0,515	0,352	0,534	$24 \cdot 10^3$	$14 \cdot 10^3$	656
	Ris.	0,561	0,727	0,872	$19 \cdot 10^3$	$36 \cdot 10^4$	78
	Var.	9,02%	106,5%	63,13%	19,2%	73,48%	88,06%

Tabella 4.11: Funzione Obiettivo con pesi (RME e RRMSE)

		RME			RRMSE		
		Corrispondenza rispetto a:			Corrispondenza rispetto a:		
		Dom. reale	conteggi	velocità	Dom. reale	conteggi	velocità
2	Rif.	0,478	0,269	0,440	0,742	0,485	0,638
	Ris.	0,504	0,124	0,153	0,806	0,325	0,250
	Var.	-5,40%	54,01%	65,13%	-8,7%	32,92%	60,82%
2 con pesi	Rif.	0,478	0,269	0,440	0,742	0,485	0,638
	Ris.	0,386	0,098	0,117	0,491	0,234	0,166
	Var.	19,15%	63,54%	73,37%	33,8%	51,69%	74,03%
6	Rif.	0,538	0,392	0,497	0,882	0,620	0,659
	Ris.	0,528	0,384	0,424	0,899	0,615	0,582
	Var.	1,85%	2,05%	14,80%	-1,9%	0,75%	11,73%
6 con pesi	Rif.	0,538	0,392	0,497	0,882	0,620	0,659
	Ris.	0,486	0,143	0,187	0,793	0,319	0,228
	Var.	9,70%	63,44%	62,50%	10,1%	48,50%	65,45%

Nel caso di una matrice target con più congestione rispetto alle condizioni effettive della rete (Test 8), ci si può trovare di fronte ad una limitazione della procedura di correzione causata dal fenomeno della congestione; infatti qualora la congestione sia particolarmente rilevante, a causa dei vincoli di capacità della rete (implicitamente rispettati nel processo di

simulazione), i flussi sono costanti ed indipendenti dal livello di domanda.

In tal caso, differenti valori di domanda possono generare la stessa corrispondenza in termini di flussi all'interno della funzione obiettivo (Domanda di riferimento e domanda risultante – Test 8, Tabella 4.12); solo le velocità agiscono in tal caso come uniche variabili sensibili (Figura 4.9), ma questa informazione non può essere da sola sufficiente per stimare la domanda “reale”.

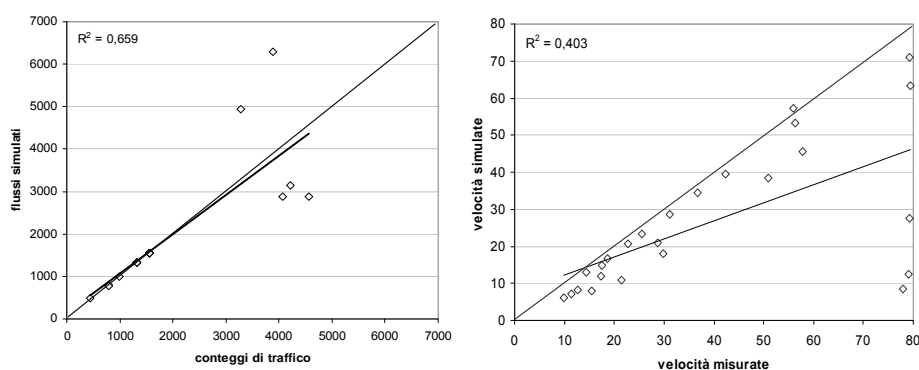


Figura 4.9: Corrispondenza in termini di flussi e velocità senza vincolo di generazione

Tabella 4.12: Introduzione del vincolo di generazione (R^2 e MSE)

		R^2			MSE		
		Corrispondenza rispetto a:			Corrispondenza rispetto a:		
		Dom. reale	conteggi	velocità	Dom. reale	conteggi	velocità
8	Rif.	0,948	0,629	0,382	$31 \cdot 10^5$	$62 \cdot 10^4$	559
	Ris.	0,926	0,922	0,998	$25 \cdot 10^4$	$99 \cdot 10^3$	2
	Var.	-2,4%	46,7%	161,2%	92,2%	83,9%	99,6%
8 senza vincolo di generazione	Rif.	0,948	0,629	0,382	$31 \cdot 10^5$	$62 \cdot 10^4$	559
	Ris.	0,947	0,659	0,403	$32 \cdot 10^5$	$55 \cdot 10^4$	523
	Var.	-0,11%	4,78%	5,45%	2,11%	-10,5%	-6,49%

Tabella 4.13: Introduzione del vincolo di generazione (RME e RRMSE)

		RME			RRMSE		
		Corrispondenza rispetto a:			Corrispondenza rispetto a:		
		Dom. reale	conteggi	velocità	Dom. reale	conteggi	velocità
8	Rif.	0,649	0,184	0,346	1,003	0,419	0,608
	Ris.	0,174	0,070	0,025	0,281	0,168	0,038
	Var.	73,2%	62,2%	92,7%	72%	59,9%	93,7%
8 senza vincolo di generazione	Rif.	0,649	0,184	0,346	1,003	0,419	0,608
	Ris.	0,644	0,172	0,324	1,014	0,396	0,588
	Var.	-0,79%	-6,46%	-6,26%	1,05%	-5,41%	-3,30%

In questo caso il vincolo di generazione come limite superiore permette di superare il povero contributo fornito dai flussi, riportando la domanda al suo corretto valore: esso ha dimostrato la sua capacità di ridurre il livello complessivo di domanda, con ottimi risultati (Tabella 4.12, Tabella 4.13) per ogni termine della funzione obiettivo.

5. Applicazione rete di Calgary

L'SPSA modificato e precedentemente applicato su di una rete test (SPSA AD-PI), viene di seguito utilizzato per correggere la domanda O-D in un contesto reale: una sottoarea della città di Calgary in Canada.

La rete è rappresentata da 734 archi, 221 nodi e 77 centroidi (Figura 5.1).

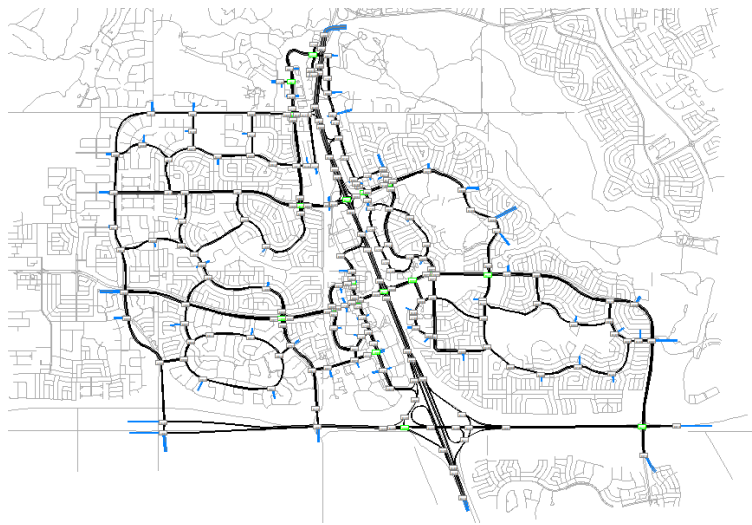


Figura 5.1: Rete di Calgary in Dynameq

Sono stati condotti numerosi esperimenti di laboratorio considerando una domanda reale nota (pari complessivamente a 12.508 veic/h) ed un numero complessivo di sezioni monitorate pari a 116 lungo il corridoio principale della rete (38 misure d'arco e 78 misure relative a manovre).

I valori osservati sono stati generati da una assegnazione dinamica all'equilibrio della matrice di domanda "reale" tramite Dynameq, mentre differenti domande target sono state generate perturbando la domanda "reale".

5.1 Esperimenti con un intervallo temporale

I primi esperimenti considerano la domanda, sia la target che la reale, distribuita uniformemente su di un'ora di simulazione (un unico intervallo temporale). Gli output (flussi e velocità) sono dunque i valori medi

riscontrati nell'intervallo di simulazione. La domanda "reale" complessiva (12.508 veic/h) è stata perturbata per ottenere differenti matrici di domanda target.

Le ipotesi adottate in merito a queste perturbazioni sono di seguito riportate:

1. sono state considerate solo le principali coppie O-D, così da coinvolgere circa il 45% della domanda complessiva "reale";
2. sono stati adottati due range di perturbazione:
 - a. Range 15÷25%;
 - b. Range 25÷35%;
3. per i differenti range sono state applicate quattro tipologie di perturbazione controllata (così da ottenere un complessivo di 8 differenti matrici di domanda target):
 - a. una perturbazione random delle principali coppie O-D (domanda random);
 - b. le principali coppie O-D perturbate tutte in eccesso (domanda alta);
 - c. le principali coppie O-D perturbate tutte in difetto (domanda bassa);
 - d. la metà delle principali coppie O-D perturbate in eccesso e l'ulteriore metà in difetto.

Tutti gli esperimenti sono stati effettuati utilizzando il vincolo di generazione e successivamente senza, per testarne l'effettiva validità nel processo di correzione O-D.

Come fatto in precedenza, anche in questo caso sono stati utilizzati gli indicatori statistici RME e R^2 per valutare i risultati. Tali indicatori sono stati inoltre accompagnati dal valore del miglioramento percentuale ottenuto in termini di funzione obiettivo.

L'analisi dei risultati con l'applicazione del vincolo di generazione (

Tabella 5.3 -

Tabella 5.4) per i differenti range di perturbazione confermano l'efficacia della procedura adottata: in tutti i casi la funzione obiettivo decresce da un valore minimo del 28% ad un valore massimo del 97%.

Questa riduzione è portata, in misura differente, dalla riduzione dei differenti termini presenti nella funzione obiettivo. Il termine relativo ai flussi d'arco è in generale più sensibile del termine relativo alle velocità, poiché può variare in un intervallo maggiore di valori. Di fatti, mentre le velocità variano tra zero e la velocità libera (differente arco per arco, ma

non superiore agli 80 km/h nella rete di studio), i flussi hanno un limite superiore molto più elevato (sono presenti anche archi con più di due corsie per senso di marcia e capacità di tipo autostradale). La conseguenza è che nella funzione obiettivo il termine relativo alle velocità si presenta più basso all'inizio rispetto al termine relativo al flusso e così le sue variazioni sono più contenute.

La riduzione della funzione obiettivo non implica sempre un miglioramento della domanda stimata: questo è ciò che accade nel caso di domanda target bassa. I risultati in termini di RME mostrano in questo caso un lieve peggioramento della domanda (+1,7% per il range 15-25%, +5,5% per il range 25-35%). Nei casi rimanenti la procedura porta un miglioramento della domanda variabile tra il -16% ed il -96% in termini di riduzione dell'RME.

I migliori risultati sono stati ottenuti quando la matrice target eccede la reale (ovvero quando il vincolo di generazione agisce su tutte le componenti – coppie O-D – e dimostra la sua efficacia). In questo caso il vincolo di generazione non è solo un limite superiore per il valore complessivo della domanda, ma (come riportato in 4.1) permette anche di superare un limite del modello che può portare ad una sovrastima della domanda: a causa dei vincoli di capacità che devono essere rispettati all'interno del modello di simulazione, in caso di condizioni di traffico particolarmente pesanti, i flussi in uscita dagli archi si mantengono costanti ed indipendenti dal livello di domanda. Così, differenti matrici caratterizzate da un'elevata domanda, eccedente la capacità, possono generare la stessa corrispondenza in termini di flusso d'arco all'interno della funzione obiettivo.

Tabella 5.1: Valori iniziali della funzione obiettivo e dell'RME - Range di perturbazione: 15÷25%

		Valori iniziali								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 15÷25%	Random	14,27	8,50	5,77	0,209	0,076	0,0044	0,887	0,995	0,988
	Domanda alta	107,22	104,45	2,77	0,209	0,136	0,0055	0,998	0,992	0,982
	Domanda bassa	41,26	39,19	2,07	0,195	0,128	0,0036	0,996	0,996	0,992
	Metà e metà	139,90	10,37	128,53	0,209	0,072	0,0061	0,874	0,995	0,923

Tabella 5.2: Valori iniziali della funzione obiettivo e dell'RME - Range di perturbazione: 25÷35%

		Valori iniziali								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Random	11,05	9,52	1,53	0,302	0,147	0,0038	0,743	0,987	0,994
	Domanda alta	203,56	73,73	129,84	0,302	0,190	0,0095	0,999	0,985	0,878
	Domanda bassa	20,10	11,47	3,16	0,291	0,184	0,0060	0,994	0,989	0,975
	Metà e metà	20,57	18,45	2,11	0,303	0,119	0,0044	0,896	0,989	0,991

Tabella 5.3: Risultati con vincolo di generazione - Range di perturbazione: 15÷25%

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 15÷25%	Random	-55	-52	-70	-45	+14	-4,6	+6,8	+0,1	+0,4
	Domanda alta	-95	-97	-36	-64	-66	-18	-2	+0,7	+0,5
	Domanda bassa	-79	-85	-26	+1,7	-1,4	-17	-4,6	-0,1	+0,2
	Metà e metà	-94	-61	-98	-21	+43	-25	+6	-0,2	+6,8

Tabella 5.4: Risultati con vincolo di generazione - Range di perturbazione: 25÷35%

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Random	-28	-40	-29	-16	+0,2	+3,3	+9,3	+0,2	-0,1
	Domanda alta	-97	-95	-99	-96	-82	-63	+0,1	+1,5	+13
	Domanda bassa	-52	-60	-20	+5,5	-8,4	-13	-11	+0,2	+1,2
	Metà e metà	-58	-67	-21	-30	-10	-23	+3,1	+0,02	-0,05

Tutti gli esperimenti sono stati ripetuti senza l'adozione del vincolo di generazione (Tabella 5.5-Tabella 5.6): in questi casi i miglioramenti in termini di funzione obiettivo sono minori, così come i miglioramenti sulla domanda stimata. Quando la matrice target è perturbata in eccesso rispetto alla reale (domanda alta), ovvero quando l'algoritmo deve ridurre

il quantitativo complessivo di domanda, la procedura senza vincolo di generazione si dimostra non efficace (peggioramento della domanda stimata).

Questo accade perché, come detto in precedenza, il termine relativo al flusso d'arco non è sensibile alle variazioni di domanda quando la congestione è elevata, così differenti livelli di domanda producono la stessa corrispondenza in termini di flussi.

E' interessante osservare cosa accade nel caso di applicazione della procedura senza vincolo di generazione, qualora la matrice target sia minore della reale (domanda bassa): in questo caso i risultati mostrano che la mancanza del vincolo comporta un miglioramento della domanda stimata.

Tabella 5.5: Risultati senza vincolo di generazione - Range di perturbazione: 15÷25%

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 15÷25%	Random	-44	-32	-78	-3,7	+10	-28	-1,5	-0,04	+0,8
	Domanda alta	-87	-92	-23	+63	-20	-27	-12	-0,01	+1
	Domanda bassa	-80	-84	-15	-3,3	-12	-3,1	-0,8	+0,07	-0,1
	Metà e metà	-94	-48	-98	-1,8	+6,1	-37	-0,2	-0,04	+7,7

Tabella 5.6: Risultati senza vincolo di generazione - Range di perturbazione: 25÷35%

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Random	-47	-61	+10	-6,9	-11	+8,5	+6,6	+0,2	-0,1
	Domanda alta	-90	-81	-98	+15	-34	-57	-8,9	+0,05	+13
	Domanda bassa	-38	-26	-75	-2,8	-4,9	-12	-6,8	+0,1	+1
	Metà e metà	-47	-61	+10	-0,8	-9,9	-0,7	-0,9	-0,01	+0,1

Sono state condotte ulteriori prove con le stesse condizioni iniziali (domanda bassa): i risultati confermano che il vincolo di generazione, anche se agisce raramente in tal caso, influenza il processo di ricerca e

può impedire all'algoritmo di incrementare il livello complessivo di domanda.

Questo accade perché, mentre in caso di domanda alta il vincolo di generazione lavora su tutte le origini e quindi su tutte le coppie O-D perturbate, negli altri casi è più probabile che vada ad agire solo su alcune di esse, lasciando le altre immutate e pregiudicando così la bontà della domanda risultante.

Per ottenere un buon risultato in queste condizioni, è necessario lavorare maggiormente sui singoli termini della funzione obiettivo. Per aggiungere maggiori informazioni alla procedura di ricerca, sia i flussi che le velocità devono lavorare in maniera efficiente; ma il termine relativo alle velocità, come riportato in precedenza, è di entità minore del termine relativo ai flussi d'arco a causa della limitatezza del suo intervallo di variazione.

La soluzione è in questo caso da trovare nella normalizzazione dei due termini di flusso e velocità al fine di bilanciare i loro contributi.

La normalizzazione può essere condotta valutando i pesi esterni dopo la prima iterazione dell'algoritmo (ovvero valutando i termini della funzione obiettivo in corrispondenza della domanda target). In tal caso il miglioramento può essere rilevante sia sui flussi che sulle velocità, con un conseguente miglioramento della domanda stimata.

Questo è proprio ciò che accade per la prova riportata in Tabella 5.7 (domanda bassa, entrambi i range di perturbazione, algoritmo di stima che include il vincolo di generazione): i risultati sono migliori di quelli ottenuti senza normalizzazione.

L'algoritmo può lavorare di più sulle velocità (variazioni del termine della funzione obiettivo relativo alle velocità anche del -75%) e può migliorare la stima della domanda anche in presenza del vincolo di generazione.

Tabella 5.7: Risultati - Normalizzazione dei termini della funzione obiettivo

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 15÷25%	Domanda bassa	-56	-77	-38	-8,8	+9,8	-4,6	-3,6	-0,3	-0,2
Perturbazione: 25÷35%	Domanda bassa	-52	-40	-75	-17	-9,3	-18	-7,3	-0,1	+1,3

5.2 Esperimenti con quattro intervalli temporali

Ulteriori esperimenti sono stati condotti considerando la domanda suddivisa in 4 intervalli temporali da 15 minuti l'uno. La matrice "reale" è stata distribuita per i differenti intervalli secondo il profilo mostrato in Figura 5.2.

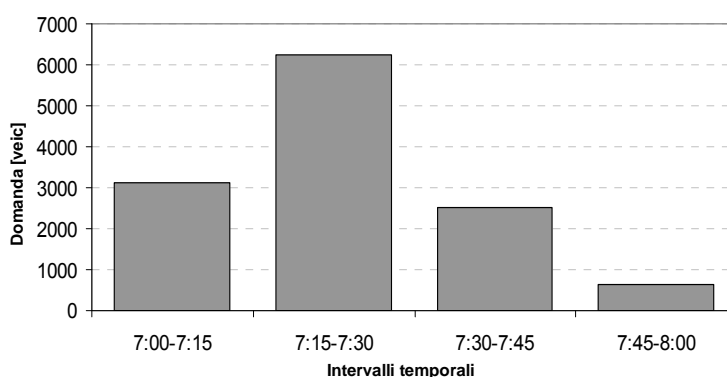


Figura 5.2: Andamento temporale della domanda "reale"

Gli output della simulazione (flussi e velocità) sono in questo caso ottenuti come valori medi per ogni intervallo temporale. Sotto queste ipotesi, il processo di correzione è stato applicato per differenti matrici di domanda di partenza ottenute, come fatto nel caso di un solo intervallo temporale, considerando solo le principali coppie O-D ed effettuando 4 tipologie di perturbazione controllata (domanda alta, bassa, random, metà e metà), ma considerando un solo range di perturbazione del 25÷35%.

Dall'analisi dei risultati con l'applicazione del vincolo di generazione (Tabella 5.9) viene dimostrato anche in questo caso come l'algoritmo sia in grado di "lavorare" rispetto al valore iniziale della funzione obiettivo, andando a ridurne il valore di ben il 99% (caso domanda alta). Inoltre in tale applicazione la riduzione suddetta non scende al di sotto del 50% (mentre nell'applicazione alla rete di Calgary con un unico intervallo temporale si ottenevano anche valori più contenuti, come il -28% del caso random a parità di range di perturbazione).

Tabella 5.8: Valori iniziali della funzione obiettivo e dell'RME (Calgary 4 intervalli temporali)

		Valori iniziali								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Random	4.655	3.337	1.318	0,299	0,160	0,148	0,839	0,982	0,8
	Domanda alta	74.211	72.028	2.183	0,297	0,263	0,181	0,999	0,936	0,715
	Domanda bassa	2.522	1.825	697	0,303	0,238	0,155	0,996	0,962	0,799
	Metà e metà	3.643	2.738	905	0,296	0,196	0,150	0,896	0,977	0,811

Tabella 5.9: Risultati con vincolo di generazione (Calgary 4 intervalli temporali)

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Random	-85	-89	-78	-7,7	-4,3	-3,1	+1,8	-0,05	+0,2
	Domanda alta	-99	-99	-96	-10	-50	-35	-8,1	+5,5	+23
	Domanda bassa	-55	-57	-58	+19	-4,4	-3,6	-9,3	-1	+1,1
	Metà e metà	-69	-68	-75	+17	-15	+1,6	-8,8	+0,6	-2,2

Anche qui non è detto che la riduzione della funzione obiettivo comporti sempre un miglioramento della domanda stimata. In particolare solo per riduzioni della funzione obiettivo superiori all'80% si è in grado di apprezzare un effettivo miglioramento della domanda, cosa che si verifica solo in due dei quattro casi analizzati (random e domanda alta). Per di più nel caso di matrice target perturbata in eccesso, si evidenzia un'ottima corrispondenza tra flussi/velocità simulati/e e misurati/e.

Nei due casi rimanenti (domanda bassa e metà e metà) è evidente il peggioramento ottenuto sulla domanda che trova corrispondenza negli scarsi risultati ottenuti su flussi e velocità.

D'altra parte nel caso di un unico intervallo temporale, si era già ottenuto per una delle precedenti ipotesi (domanda bassa) un peggioramento della domanda (variazione di RME pari ad +1,7% per il range 15-25% ed a +5,5% per il range 25-35%); aumentando la dimensione temporale del problema sono state incrementate le variabili da stimare (di quattro volte), mentre sono rimaste dello stesso numero le sezioni stradali da cui ricavare l'informazione circa lo stato del deflusso veicolare (informazioni di flusso e velocità).

Per cercare anche in questo caso di migliorare i risultati ottenuti, si è pensato di lavorare maggiormente sui singoli termini della funzione obiettivo.

Quindi, senza considerare se la matrice di partenza implicasse o meno un maggior quantitativo di domanda rispetto a quella reale, il processo di correzione è stato ripetuto normalizzando i contributi di flusso e velocità nella funzione obiettivo. Anche questa volta la normalizzazione è stata effettuata calcolando i pesi esterni dopo la prima iterazione dell'algoritmo.

Tabella 5.10: Risultati con vincolo di generazione e normalizzazione (Calgary 4 intervalli temporali)

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Random	-83	-87	-81	-9,8	-0,4	-6,4	+2,3	+0,02	+1,1
	Domanda alta	-99	-99	-87	-78	-58	-29	-0,4	+5,8	+17
	Domanda bassa	-57	-66	-51	+8,0	-16	-3,7	-11	+0,6	+1,5
	Metà metà	-73	-72	-75	+8,7	-5,2	-14	-2	+0,1	+6,1

I risultati migliorano notevolmente a seguito dell'introduzione dei pesi esterni (Tabella 5.10). Anche se negli ultimi due casi (domanda bassa e metà e metà) non si riesce ancora a raggiungere il risultato cercato in relazione alla domanda, l'incremento della statistica RME si mostra più contenuto; così i relativi miglioramenti in termini di flussi e di velocità diventano più forti.

Quest'ultimo esperimento sottolinea l'importanza della fase di normalizzazione al fine di ottenere una soluzione "ottimale": la normalizzazione, e quindi la corretta specificazione della funzione obiettivo, permette alla procedura di tener conto di tutte le informazioni a disposizione nel modo migliore.

Il risultato raggiunto in termini di corrispondenza tra domanda stimata e domanda reale è riportato a titolo di esempio per il caso migliore (domanda alta) nello scattergramm di Figura 5.3.

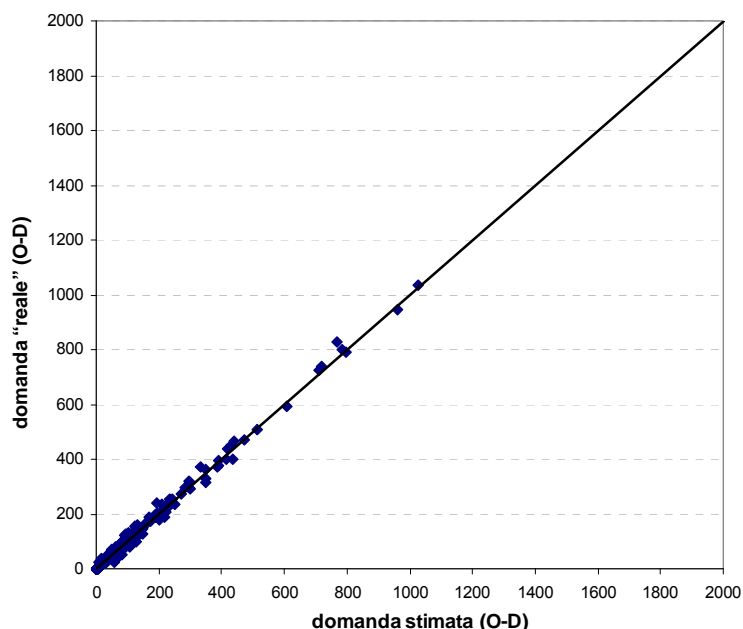


Figura 5.3: Confronto tra domanda stimata e domanda reale

5.2.1 Influenza del tipo e della numerosità delle informazioni sulla procedura di correzione

Gli esperimenti condotti nei precedenti paragrafi sulla rete di Calgary hanno utilizzato come input della procedura di correzione, a parte la domanda target, una serie di misure di traffico desunte da 38 archi stradali e 13 intersezioni monitorate, entrambi disposti lungo il corridoio principale della rete. Tali sezioni di monitoraggio forniscono un complessivo di 38 misure d'arco e 78 misure relative a manovre. Per ogni arco e per ogni manovra sono a disposizione le informazioni di flusso e velocità, ottenendo così un totale di 116 misure di flusso e 116 misure di velocità per ogni intervallo temporale.

Ovviamente nel caso in cui il periodo di analisi venga considerato come un unico intervallo temporale, i dati relativi a flussi e velocità saranno pari ai valori medi ottenuti in simulazione, mentre se il periodo di analisi viene suddiviso in più finestre temporali (cfr. paragrafo 5.2), si otterrà

un'informazione più accurata circa la variazione delle misure stesse nel tempo.

In Dynameq sugli archi viene definita una curva di deflusso di tipo triangolare (Figura 3.12). Al contrario per le intersezioni non semaforizzate viene utilizzato un modello di *gap-acceptance* (paragrafo 3.6.4); quest'ultimo è ripreso anche per regolare eventuali manovre in conflitto nel caso di intersezioni semaforizzate.

Mentre dunque il deflusso sull'arco è sufficientemente chiaro ed è quindi immediato il legame esistente tra flusso e velocità, alle intersezioni esiste una fonte di stocasticità che deriva da quando il veicolo raggiunge l'intersezione stessa, dai volumi con cui entra in conflitto, nonché dalla lunghezza degli archi di manovra.

A tal fine si è voluto analizzare l'impatto delle differenti forme di misure e della loro quantità sul processo di correzione, per capire se ognuna di esse sia effettivamente in grado di aggiungere informazione sul valore dei flussi OD.

Dai risultati delle diverse applicazioni alla rete di Calgary, si nota, anche nel caso in cui venga ottenuta un'ottima soluzione in termini di domanda stimata (Tabella 5.10 – ipotesi di domanda alta), come la corrispondenza tra flussi simulati e flussi rilevati sia ottimale (Figura 5.4), mentre diversi punti possono essere individuati nello scattergramm delle velocità lontani dalla bisettrice del quadrante (retta $y=x$, Figura 5.5).

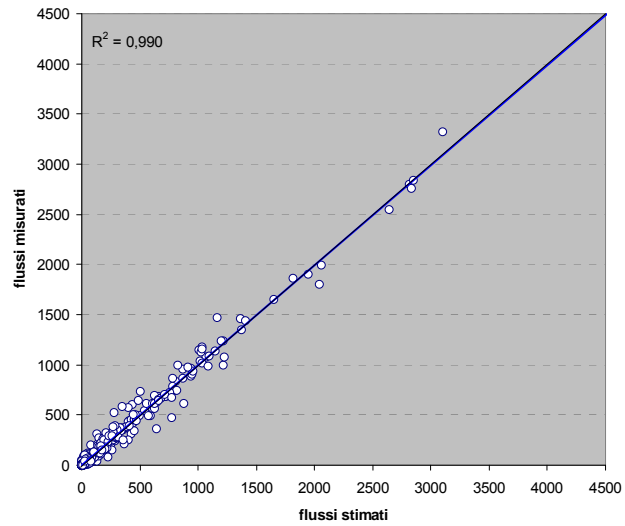


Figura 5.4: Scattergramm flussi rilevati/flussi misurati (Tabella 5.10 ipotesi di domanda alta)

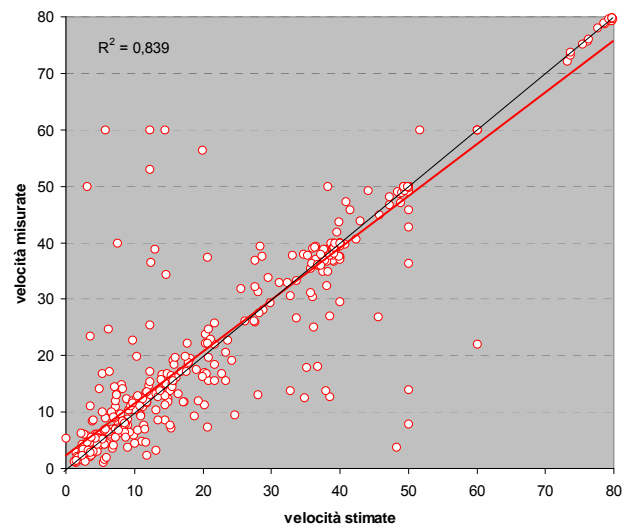


Figura 5.5: Scattergramm velocità stimate/velocità misurate (Tabella 5.10 ipotesi di domanda alta)

Volendo risalire a quale tipologia di informazione fanno riferimento i punti in questione, si scopre che tali punti sono ottenuti per flussi di entità modesta e principalmente relativi a manovre di svolta a sinistra.

Per capire dunque il perché a tali valori di flussi possano fare riferimento valori così differenti di velocità di manovra è stata condotta un'analisi circa la relazione tra le variabili flusso e velocità alle intersezioni.

I risultati delle numerose simulazioni effettuate sono stati raccolti e riportati nel grafico di Figura 5.6, ordinati per valori di flusso crescente. Nel grafico sono presenti tutte le manovre, siano esse di attraversamento, di svolta a destra o di svolta a sinistra.

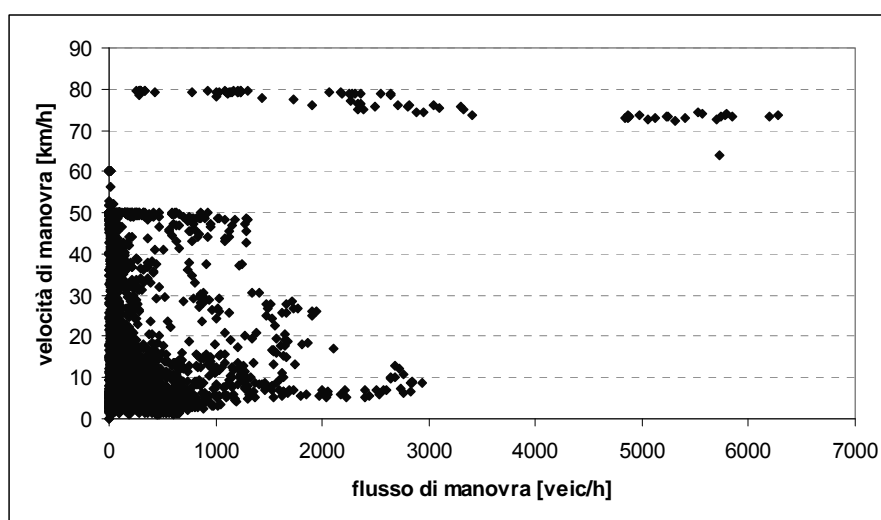


Figura 5.6: Flussi e velocità di manovra alle intersezioni

Si nota la presenza di una forte concentrazione dei valori per un flusso veicolare inferiore ai 1.000 veic/h. La presenza di una serie di punti con velocità prossime a 70-80 km/h e flussi superiori ai 4.000 veic/h fa riferimento ad un circoscritto numero di manovre di attraversamento in corrispondenza di una viabilità a tre corsie per senso di marcia con una capacità per corsia superiore ai 2.000 veic/h (ambito autostradale). Analogamente l'insieme di punti ad 80 km/h per flussi inferiori ai 4.000 veic/h sono da riferire a dati raccolti in differenti intervalli temporali sempre per la stessa manovra di attraversamento, localizzata lungo una viabilità che permette il mantenimento di elevate velocità. Stesso discorso va fatto per l'insieme di punti localizzati in corrispondenza del valore

della velocità di 50 km/h; in quest'ultimo caso però alle manovre di attraversamento si uniscono manovre di svolta a destra.

Cosa succede invece per le manovre di svolta a sinistra, ovvero per quelle manovre da cui dipendono prettamente i punti lontani dalla bisettrice nello scattergramm delle velocità?

In Figura 5.7 sono stati selezionati per l'appunto solo tali manovre, da cui si deduce un andamento delle velocità in funzione dei flussi di tipo esponenziale negativa.

Come si può notare la variabilità dei valori assunti dalle velocità aumenta man mano che ci si avvicina a valori dei flussi prossimi allo zero.

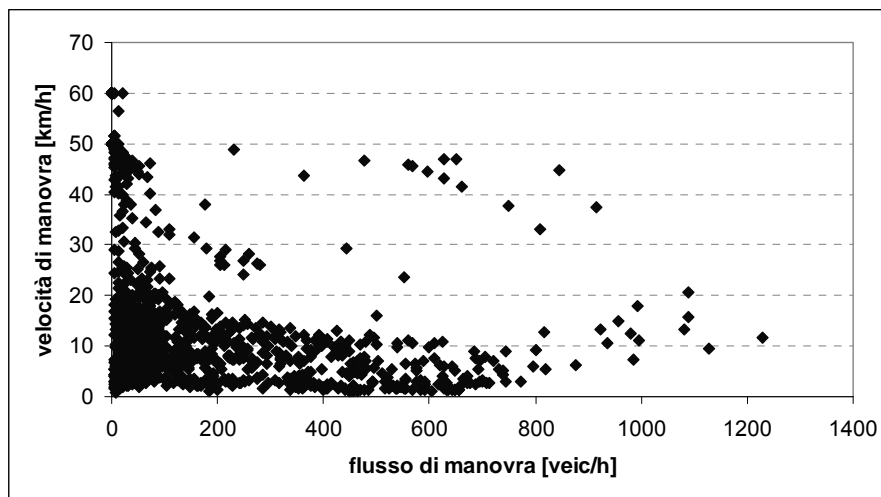


Figura 5.7: Flussi e velocità delle manovre di svolta a sinistra alle intersezioni

Con flussi così bassi, i tempi di percorrenza sono pressochè random. Nel caso di intersezioni semaforizzate i veicoli che arrivano all'intersezione semaforizzata poco prima che stia per scattare il verde, avranno un ritardo minore, mentre i veicoli che arrivano all'inizio del rosso subiranno un ritardo maggiore. Analogamente, nel caso di intersezioni non semaforizzate, a seconda dell'istante in cui il veicolo arriva al nodo, troverà differenti valori dei flussi di conflitto e di gap utilizzabili. Ciò, combinato al fatto che gli archi di manovra possono essere particolarmente corti, porta ad avere un impatto piuttosto significativo sui ritardi complessivi (e quindi sulle velocità).

Per evitare che tale variabilità del dato “velocità” vada ad influire negativamente nel processo di correzione, si è ipotizzato di conoscere l’informazione di flusso sia per gli archi che per le manovre, mentre che le velocità siano note solo sugli archi. Tale ipotesi è d’altronde molto più realistica che considerare anche le velocità di manovra: difatti, mentre i flussi di manovra sono misurabili e rimangono molto importanti ai fini del processo di correzione della domanda perché forniscono dati sulle scelte di percorso, le velocità di manovra sono di difficile misurazione e la movimentazione della congestione sulla rete è sufficientemente spiegata dalle velocità sugli archi.

Prendendo dunque come prova di riferimento quella effettuata nell’ipotesi di domanda perturbata per metà in eccesso e per metà in difetto (Tabella 5.8 – ipotesi domanda metà in eccesso e metà in difetto), è stato ripetuto il processo di correzione della domanda eliminando l’informazione relativa alle velocità di manovra.

Le prove sono state ripetute sia senza pesi che normalizzando i due termini di flusso e velocità.

Tabella 5.11: Risultati senza velocità di manovra (Calgary 4 intervalli temporali)

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 15÷25%	Metà e metà	-78	-80	-8	+45	-7,5	-8,8	-16	+0,3	+0,03
Perturbazione: 25÷35%	Metà e metà (con normalizzazione)	-54	-56	-54	+5,7	-4,2	-24	-2,5	+0,4	+3,9

Effettuando un confronto a parità di condizioni, si può notare come nel caso senza pesi, l’aver eliminato l’informazione delle velocità di manovra (Tabella 5.11 – Metà e metà) porta da un lato ad un peggioramento della domanda stimata rispetto al caso con velocità di manovra (Tabella 5.9 – Metà e metà), dall’altro una corrispondenza migliore rispetto ai conteggi (siano essi flussi o velocità). Ma perché un peggioramento così forte sulla domanda stimata in termini di RME? L’aver ridotto il numero di informazioni, in particolare in questo caso le velocità di manovra, porta sicuramente a disporre di minori informazioni circa lo stato del deflusso sulla rete; questo fattore è comunque negativo nel processo di correzione, perché riduce la numerosità del campione di dati da cui ricavare informazioni sulla domanda. D’altra parte l’eccessiva variabilità delle

velocità di manovra, per non dimenticare la loro effettiva difficoltà di rilievo, porta ad ottenere scarse corrispondenze rispetto ai dati misurati ed inficia la qualità della stima.

Normalizzando i termini di flussi e velocità, si riesce anche in questo caso a sfruttare al massimo le potenzialità delle informazioni a disposizione, così gli ottimi risultati ottenuti in termini di corrispondenza rispetto alle misure sulla rete, riducono il peggioramento ottenuto sulla domanda stimata (solo +5,7% in termini di variazione dell'RME). Non si riesce ad ottenere ancora un miglioramento sulla domanda e ciò probabilmente in virtù della quantità di informazioni utilizzate, che possono non risultare sufficienti.

Per verificare se è proprio la numerosità delle informazioni utilizzate la motivazione che porta a non ottenere un miglioramento della domanda durante la procedura di correzione, il numero di rilievi è stato incrementato aggiungendo ulteriori 40 sezioni monitorate. Sono state così ottenute un totale di 78 manovre monitorate (per le quali viene utilizzato, secondo quanto spiegato precedentemente, la sola informazione del flusso di manovra) e 78 archi monitorati (per i quali si dispone invece del dato relativo al flusso ed alla velocità).

La procedura di correzione è stata quindi ripetuta andando direttamente a normalizzare i termini di flusso e velocità e considerando queste informazioni aggiuntive:

Tabella 5.12: Risultati senza velocità di manovra con incremento sezioni d'arco (Calgary 4 intervalli temporali)

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Metà e metà (con normalizzazione)	-73	-67	-82	-0,5	-12	-22	-1,3	+0,3	+6,6

I risultati (Tabella 5.12) confermano quanto ipotizzato: la quantità e la qualità delle informazioni utilizzate sono determinanti nel processo di correzione per poter ottenere un miglioramento sulla domanda.

Certo il miglioramento riscontrato appare ancora contenuto (riduzione dell'RME dello 0,5%) e non raggiunge la qualità ottenuta nel caso di domanda in eccesso. Le motivazioni che conducono a tale risultato verranno studiate nei paragrafi seguenti.

5.2.2 Intervallo di simulazione e dimensione della domanda

Nel processo di assegnazione dinamica assumono particolare importanza due fattori che nel caso di simulazione statica risultano meno rilevanti per l'ipotesi di stazionarietà intraperiodale:

- l'intervallo di simulazione inteso come il periodo temporale nel quale si vogliono rappresentare i fenomeni di traffico in questione;
- la dimensione della domanda intesa come dimensione quantitativa e come conseguente dimensione temporale (tempo necessario a caricare la domanda sulla rete e tempo necessario affinché tutti gli utenti raggiungano la rispettiva destinazione).

Nel caso di assegnazione statica, se ad esempio si considera l'ora di punta come intervallo di simulazione, la rispettiva domanda fa esattamente riferimento a tale intervallo e viene completamente simulata nell'intervallo stesso.

Differentemente nel caso di assegnazione dinamica, poiché si perde l'ipotesi di stazionarietà intraperiodale, la domanda verrà completamente generata nell'intervallo di simulazione, ma non è detto che ogni singolo spostamento venga completato nell'intervallo di simulazione stesso, a causa dei fenomeni di congestione e dei conseguenti tempi di percorrenza che si vengono a determinare sulla rete. Questo fattore equivale a dire che non tutta la domanda emessa raggiunge obbligatoriamente la propria destinazione all'interno dell'intervallo di simulazione.

Tornando alla suddetta ora di punta, potrebbero dunque esistere spostamenti che impiegano più di un'ora a raggiungere la propria destinazione, così come potrebbero esserci spostamenti generati nell'ultimo quarto d'ora e che a maggior ragione, a seconda della propria lunghezza, vengono simulati solo in parte, mentre la restante fase dello spostamento ricade solo dopo il periodo di simulazione prefissato.

Quando a questo discorso si affianca il problema di correzione della domanda, date le misure di traffico rilevate nel solo periodo di simulazione, nasce un interrogativo fondamentale: quale domanda si sta cercando di ottenere? La domanda che nella realtà genera le condizioni di traffico che si vogliono ribattere o la domanda che rispecchia le condizioni di capacità della rete e contemporaneamente le condizioni di traffico note? Se difatti la domanda da stimare impiega un periodo di

tempo per essere completamente simulata (dall'origine sino alla destinazione per ogni spostamento) superiore all'intervallo di simulazione, i volumi di traffico e le relative velocità con cui ci si va a confrontare non sono altro che lo "specchio" delle condizioni di capacità sugli archi e sui nodi della rete. A questo punto la domanda che il metodo è in grado di stimare non può che essere al massimo pari alla domanda che la rete, nelle sezioni in cui essa è monitorata, è in grado di accogliere (Figura 5.8).

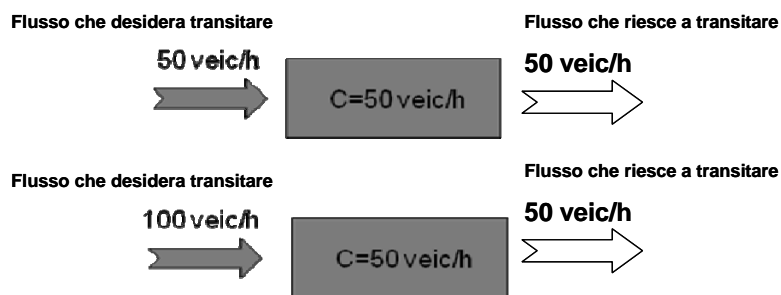


Figura 5.8: Interazione tra domanda e vincoli di capacità della rete

Questo fattore, seppur formalmente corretto perché comunque la domanda stimata andrebbe a ribattere le uniche informazioni reali a disposizione (flussi e velocità misurate), risulta limitativo per il raggiungimento della conoscenza della domanda "reale".

Per superare tale limite si è ipotizzato di seguire l'evoluzione della domanda per un periodo di simulazione maggiore dell'ora stabilita negli esperimenti precedenti ed in particolare pari al tempo necessario per lo scaricamento della domanda "reale" (75 minuti), raccogliendo informazioni alle sezioni monitorate sull'intero periodo di simulazione (sino a che l'ultimo veicolo esce dalla rete).

Si deve tener presente che per agire in tal senso sono state fatte una serie di ipotesi molto forti:

- la prima è che effettuando esperimenti di laboratorio si partiva dalla conoscenza della domanda "reale" e quindi del tempo effettivamente necessario per lo scaricamento della stessa. In condizioni normali la domanda "reale" non è ovviamente nota, essendo l'output del processo di correzione, e dunque non è noto con certezza il tempo necessario al suo scaricamento;

- in condizioni reali non potremmo effettivamente seguire l'evoluzione della domanda da stimare sino allo scaricamento, perché ad essa si andranno sempre a sommare contributi di domanda generati in intervalli temporali successivi; al massimo in condizioni reali si può parlare di riduzione graduale del livello di domanda tale da non avere situazioni di congestione sulla rete.

E' chiaro, da quanto riportato sin ora, come la dimensione quantitativa della domanda ed il correlato tempo necessario al suo scaricamento abbiano un ruolo decisivo nel limitare il processo di correzione; il processo non potrà far altro che ribattere i vincoli di capacità sulla rete se non si segue l'evoluzione della domanda sino al suo scaricamento. D'altronde la simulazione dello scaricamento della rete è qualcosa che difficilmente può ritrovare il suo corrispettivo nella realtà, perché una serie di contributi di domanda si vanno a sommare a ciò che si vuole stimare e quindi difficilmente si otterrà un'informazione di esclusiva pertinenza di quest'ultima.

Le sperimentazioni di seguito proposte, le quali includono la simulazione dello scaricamento della rete, vengono condotte con la consapevolezza di queste problematiche, ma con la contemporanea certezza di ottenere utili informazioni per la procedura di correzione.

Prima di procedere alla correzione della domanda sulla rete di Calgary tenendo conto dello scaricamento, sono state fatte alcune considerazioni relative alla funzione obiettivo definita in (3.1) e (3.2). L'errore calcolato tra valori simulati/stimati e valori reali/noti porta di fatti alcune problematiche che emergono in maniera incisiva qualora ci si trovi a confrontare valori ridotti della variabile analizzata.

Si prenda il caso riportato in Tabella 5.13, dove la differenza tra un conteggio di 1 veic/h ed un flusso simulato di 3 veic/h genera un errore del 400% (calcolato secondo quanto riportato in 3.1 e 3.2); in tal caso si registra un'amplificazione eccessiva di tale differenza che, invece, ai fini trasportistici è prettamente nulla. Al contrario, differenze che in termini trasportistici possono contare molto (un flusso di 1.500 veic/h a fronte di una valore rilevato di 2.000 veic/h), generano secondo tale formulazione della funzione obiettivo un errore ridotto (6,25%).

Tabella 5.13: Stima dell'errore per differenti valori delle variabili

Conteggio [veic/h]	Flusso simulato [veic/h]	Errore [%]
1	3	400
2.000	1.500	6,25

Se nelle prove precedenti questo problema non è stato tenuto in considerazione, in quanto sono rari i casi di valori così ridotti delle variabili in gioco, nel voler simulare lo scaricamento non è più possibile soprassedere, perché, soprattutto negli ultimi intervalli temporali corrispondenti all'uscita dalla rete dei veicoli rimasti, si avrà spesso a che fare con valori ridotti di flusso.

Per evitare dunque che la funzione obiettivo tenga principalmente conto di queste differenze irrisorie, si è pensato di adottare l'indicatore statistico RME come misura delle differenze tra valori simulati/stimati e valori reali/noti. La funzione obiettivo assume in tale caso la seguente formulazione:

$$(\mathbf{d}_1^* \dots \mathbf{d}_{n_h}^*) = \arg \min_{(\mathbf{x}_1 \dots \mathbf{x}_{n_h}) \geq 0} \left[\frac{\sum_{h=1}^{n_h} |\hat{\mathbf{d}}_h - \mathbf{x}_h|}{\sum_{h=1}^{n_h} \mathbf{x}_h} + \frac{\sum_{h=1}^{n_h} |\hat{\mathbf{v}}_h - \mathbf{v}_h|}{\sum_{h=1}^{n_h} \mathbf{v}_h} + \frac{\sum_{h=1}^{n_h} |\hat{\mathbf{s}}_h - \mathbf{s}_h|}{\sum_{h=1}^{n_h} \mathbf{s}_h} \right] \quad (5.1)$$

La procedura di ricostruzione della matrice è stata quindi applicata secondo le seguenti ipotesi:

- si è considerata la rete di Calgary, con profilo di domanda variabile (cfr. par. 5.2);
- domanda “reale” perturbata (per le coppie od principali) per metà in eccesso e per metà in difetto del 25-35% ottenendo così la domanda target (Tabella 5.8, ipotesi domanda metà e metà);
- sono state escluse le velocità di manovra dalla procedura, tenendo conto di un complessivo di 78 manovre monitorate e 78 archi monitorati (cfr. par. 5.2.1)
- la simulazione è stata condotta sino al completo scaricamento della rete raccogliendo informazioni alle sezioni ed ai nodi monitorati sino all'ultimo intervallo;

- l'indicatore statistico RME è stato adottato per la formulazione della funzione obiettivo così come riportato in (5.1).

I risultati dell'applicazione della procedura al caso con scaricamento della rete sono riportati graficamente in Figura 5.9 e Figura 5.10, nonché in Tabella 5.14; ciò che si può evidenziare è un sostanziale miglioramento della corrispondenza con conteggi e velocità misurate, unica informazione di cui si dispone nella realtà sull'effettivo stato della rete. Tali miglioramenti sono quantificati in un 33% sui flussi ed in un 37% sulle velocità (riduzione dell'indicatore statistico RME ed, in questo caso, conseguentemente dei due rispettivi termini della funzione obiettivo).

In termini di domanda, la matrice ottenuta è senz'altro una buona matrice (R^2 pari a 0,826) in grado di ribattere i dati di traffico noti, ma non è ancora la domanda "reale" ed inoltre risulta essersi allontanata da quest'ultima più della matrice target.

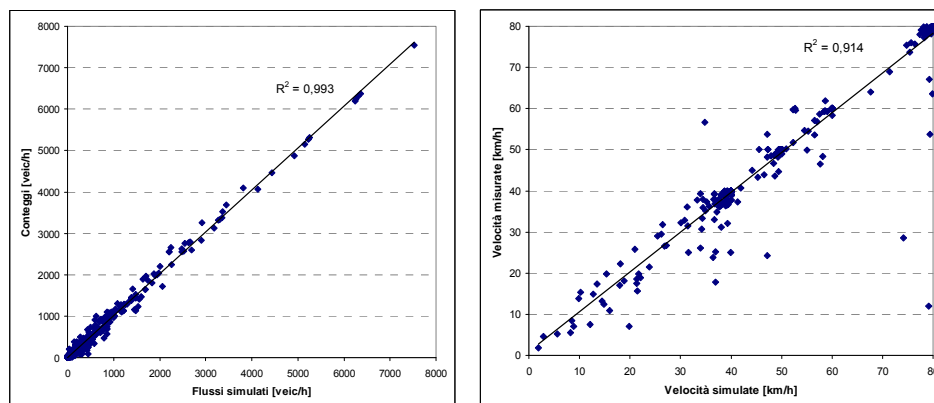


Figura 5.9: Scattergramm flussi/velocità rilevati/e vs flussi/velocità misurati/e (Calgary scaricamento)

Data la complessità del problema, è già noto come a differenti valori dei dati di traffico (flussi e velocità) possano corrispondere differenti matrici, per di più molto diverse tra loro. I risultati appena riportati ne sono una ulteriore conferma: si è ottenuta una matrice più lontana dalla "reale" di quanto lo fosse la matrice di partenza, ma in grado di restituire un'informazione in simulazione molto più prossima all'effettivo stato della rete.

Lo scaricamento ha d'altra parte permesso di raggiungere la suddetta corrispondenza in termini di flussi e velocità, risultato di gran lunga superiore a quanto ottenuto senza simulazione dello scaricamento (cfr. Tabella 5.12).

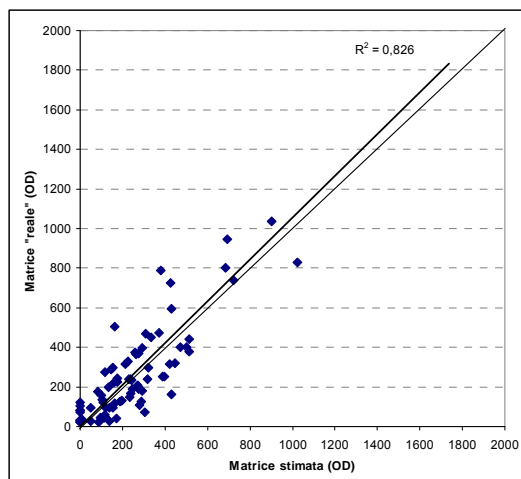


Figura 5.10: Scattergramm matrice “reale” vs matrice stimata (Calgary scaricamento)

Tabella 5.14: Risultati con scaricamento della rete (Calgary)

		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Perturbazione: 25÷35%	Metà e metà (con normalizzazione)	-31	-33	-37	+22	-33	-37	-7,8	+0,5	+8,8

5.2.3 Introduzione di vincoli più restrittivi sulla domanda

La ricostruzione effettuata nel precedente paragrafo si può considerare valida ai soli fini della simulazione, in quanto la domanda stimata è in grado di riprodurre fedelmente in simulazione le condizioni di traffico monitorate sulla rete. Tale risultato non può tuttavia considerarsi sufficiente in quanto l'informazione contenuta nella matrice non è veritiera ed, ancor più, se si volesse operare con un intervento sulla rete,

non è detto che i risultati ottenuti in simulazione andrebbero a rispecchiare gli effettivi comportamenti di mobilità.

Per superare questi limiti e per avvicinarsi all'effettiva domanda "reale" è necessario intervenire nella procedura andando ad inserire nuovi vincoli, più restrittivi, che siano in grado di aumentare le informazioni a disposizione sulla domanda stessa.

Sia nel caso della rete test che nella seguente applicazione di Calgary è stata dimostrata la validità del vincolo di generazione come limite superiore in grado di prevenire fenomeni di sovrastima della domanda che possono incorrere qualora si abbia a che fare con una rete congestionata.

Il vincolo di generazione è però in grado di indirizzare la ricerca verso una matrice "buona" solo qualora abbia l'effettiva possibilità di essere messo in gioco e soprattutto qualora riesca a lavorare contemporaneamente su tutte le zone di emissione (si confrontino in particolare i casi di ricostruzione nel caso di domanda di partenza perturbata in eccesso rispetto alla matrice reale, Figura 5.3).

Data la validità e la robustezza dell'informazione contenuta nelle generazioni da modello, è possibile pensare di modificare tale vincolo in un vincolo di uguaglianza stretta che possa lavorare sulla domanda in termini quantitativi ed in qualsiasi caso di livello di congestione rispetto allo stato reale.

Riprendendo il caso affrontato nel precedente paragrafo (con scaricamento) è stato dunque sostituito il vincolo di generazione come limite superiore con un vincolo di uguaglianza; stessa operazione (uguaglianza delle generazioni) è stata poi condotta senza effettuare lo scaricamento.

I risultati (Figura 5.11 - Figura 5.12) dimostrano la necessità di lavorare in tal senso per poter giungere ad una ricostruzione effettiva della domanda reale: solo aggiungendo l'informazione della generazione l'indicatore RME riporta un miglioramento sensibile della domanda (11% con scaricamento e 17% senza scaricamento, Tabella 5.15), senza tralasciare i miglioramenti su flussi e velocità che rimangono dello stesso ordine di grandezza di quelli ottenuti nel caso precedente (Tabella 5.14).

L'uguaglianza delle generazioni ha mantenuto il quantitativo di domanda emessa da ogni zona costante e pari al valore della generazione reale ottenuta da modello, permettendo così all'algoritmo di lavorare sulla sola distribuzione spaziale e temporale della domanda.

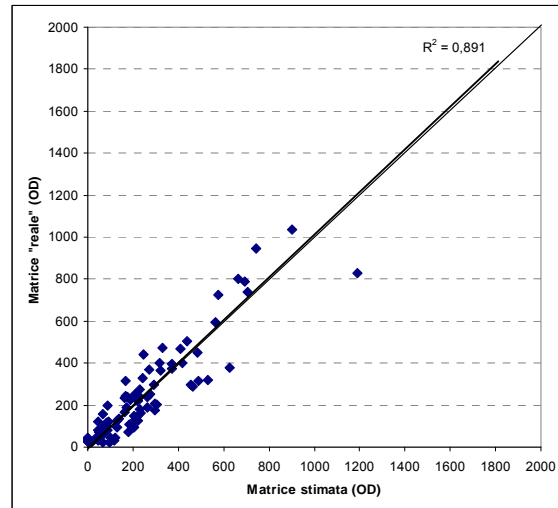


Figura 5.11: Scattergramm matrice “reale” vs matrice stimata (Calgary scaricamento con uguaglianza generazioni)

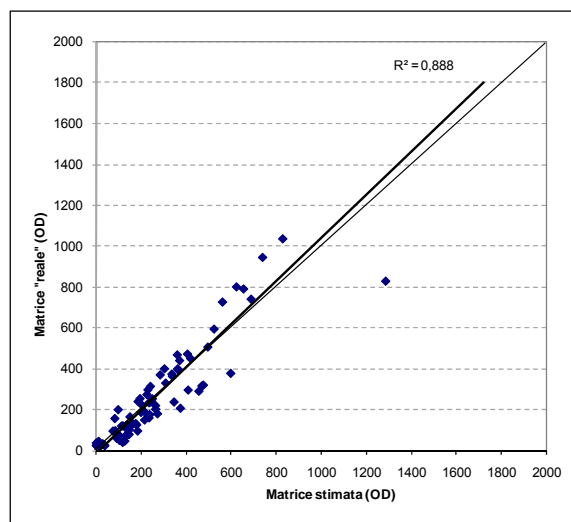


Figura 5.12: Scattergramm matrice “reale” vs matrice stimata (Calgary senza scaricamento con uguaglianza generazioni)

Tabella 5.15: Risultati con uguaglianza delle generazioni (Calgary)

Perturbazione: 25÷35%		Variazione [%]								
		F.O.			RME			R ²		
		Totale	Flussi	Velocità	Domanda	Flussi	Velocità	Domanda	Flussi	Velocità
Con scaricamento	Metà e metà (con normalizzazione)	-27	-28	-33	-11	-28	-33	-0,5	+0,3	+8,9
Senza scaricamento	Metà e metà (con normalizzazione)	-40	-38	-44	-17	-38	-44	-0,9	+0,6	+15

L'introduzione dell'uguaglianza delle generazioni rafforza, come dimostrato, la ricerca in direzione della domanda "reale"; si può allora pensare di effettuare un ulteriore passo aggiungendo altre informazioni aggregate sulla domanda ed in particolare:

- in analogia con le generazioni, informazioni circa le attrazioni da modello (teoria dell'inizio dello spostamento/interviste a destinazione);
- informazioni sulla domanda complessiva a livello di sottomatrici (dato noto tramite ad esempio l'effettuazione di indagini al cordone).

5.2.4 Dimensione del problema e tempi computazionali

Il problema di correzione/stima combinata della domanda di trasporto in dinamica intraperiodale è un problema complesso; tra le motivazioni di tale complessità risiede, come già detto nel paragrafo 3.3, l'essere un problema a larga scala, in quanto concorrono nel determinarne la dimensione sia il numero di coppie origine/destinazione che il numero di intervalli temporali in cui viene suddiviso il periodo di analisi.

Data dunque l'elevata dimensione del problema, la scelta dell'approccio da seguire sia esso simultaneo o sequenziale deriva senz'altro dal contesto applicativo a cui ci si riferisce (on-line od off-line).

Di fatti la possibilità di potersi orientare in questa ricerca verso un approccio simultaneo, formalmente più corretto, deriva proprio dall'aver supposto di lavorare in un contesto off-line.

Nel caso off-line la domanda deve essere stimata per una successiva fase pianificatoria, viceversa nel caso on-line l'aggiornamento in tempo reale dei valori dei flussi OD avviene tramite le misure di traffico disponibili mano a mano che i sistemi di rilevamento acquisiscono il dato e ciò

permette di operare in tempo reale sulla rete attraverso interventi di regolazione e controllo. Dunque in un contesto on-line i tempi computazionali diventano un vincolo essenziale del processo di stima.

Ovviamente anche nel caso off-line i tempi computazionali necessari alla procedura di stima, per quanto più lunghi, devono comunque mantenersi in un campo di variabilità sufficientemente limitato, con il fine di poter fornire quanto prima al progettista l'informazione sulla domanda per l'usuale pratica professionale.

Negli esempi applicativi riportati in letteratura, sia nel caso di approccio simultaneo che nel caso di approccio sequenziale, le reti utilizzate sono sempre di ridotte dimensioni (sezioni autostradali con rampe di ingresso/egresso – Figura 5.13, assi stradali paralleli con possibilità di interscambio – Figura 5.14) e qualora si cerchi di andare verso reti più prossime a contesti reali (Figura 5.15), i tempi computazionali salgono vertiginosamente (5 giorni per ogni iterazione).

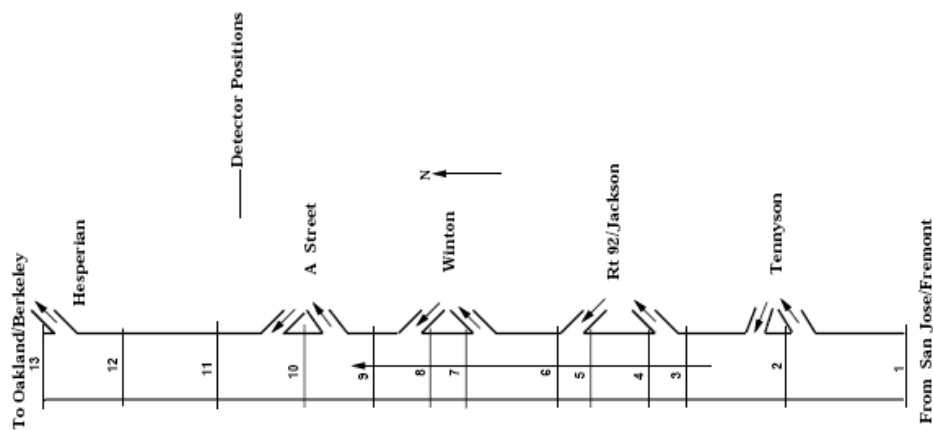


Figura 5.13: Applicazione alla Sezione autostradale I-880 California (Ashok, 1996)

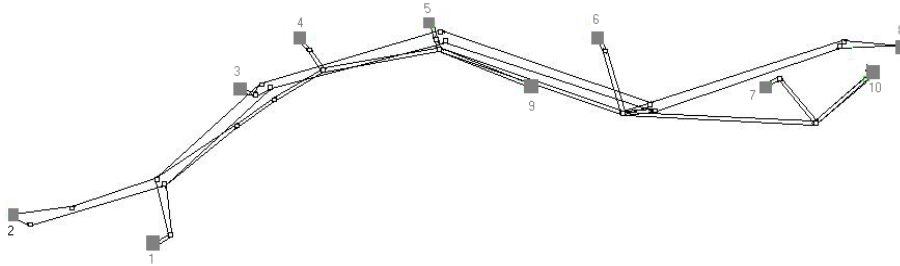


Figura 5.14: Applicazione “Waterfront corridor”, Toronto (Kattan e Abdulhai, 2005)



Figura 5.15: Applicazione alla rete di Los Angeles con 33 zone “attive” (Balakrishna, 2006)

Detto ciò, alcuni commenti devono essere fatti in merito ai tempi computazionali dell’algoritmo proposto al fine di valutarne l’efficienza. Il dispendio computazionale è riportato qui in funzione del numero di assegnazioni dinamiche (DTA) effettuate ad ogni iterazione del processo di correzione. Il numero di assegnazioni per iterazione dipende

principalmente dalla fase di calcolo del gradiente approssimato medio, poiché è richiesta una assegnazione per ognuna delle m stime del gradiente; m è funzione del numero di variabili secondo la seguente relazione:

$$m = k n_v, \quad \text{con } k < 1 \quad (5.2)$$

dove:

$$n_v = n_h \times n_{od}$$

Negli esperimenti riportati il valore di m è stato posto in prima approssimazione pari a $0,4n_v$.

Negli ultimi esperimenti con quattro intervalli temporali e dunque con un numero di variabili maggiori da stimare, la soluzione è stata trovata in massimo 15 ore (processore dual core, 3.00 GHz).

Inoltre è da dire che il metodo SPSA modificato sembra raggiungere un buon miglioramento sulla domanda con un numero relativamente basso di iterazioni (tutti i risultati riportati per la rete di Calgary sono stati ottenuti con al più 50 iterazioni).

Ciò detto, l'algoritmo SPSA modificato (SPSA AD-PI) applicato al processo di correzione della domanda, sembra essere appropriato al contesto off-line. Sicuramente un'analisi di sensitività del parametro m sarà utile per capire quanto ancora potrebbero essere ridotti i tempi computazionali.

6. Conclusioni e futuri sviluppi

Nella presente tesi si è affrontato il problema di correzione/stima combinata della domanda di trasporto attuale in dinamica intraperiodale per un contesto di tipo off-line.

Partendo dalla formulazione simultanea formalizzata da Cascetta *et al.* (1993), il problema è stato esteso andando a considerare ulteriori misure di traffico quali le velocità per poter rappresentare univocamente lo stato, congestionato o non congestionato, di ogni elemento della rete.

Sia nella formulazione che nella successiva procedura risolutiva sono stati inseriti aspetti innovativi con l'intento di migliorare la soluzione e di ridurre i tempi computazionali.

Per quanto attiene la formulazione, l'elemento più innovativo consiste nell'introduzione di un nuovo vincolo relativo alla generazione complessiva di ogni singola zona; questa tipologia di informazione risulta facilmente disponibile (attraverso l'applicazione dei modelli matematici dell'Ingegneria dei Trasporti) ed inoltre è senz'altro considerata altamente affidabile. Ma il vincolo di generazione non è solo un elemento aggiuntivo in grado di produrre ulteriori informazioni sulla domanda: difatti si dimostra essere fondamentale nel prevenire la sovrastima della domanda che può verificarsi in caso di reti particolarmente congestionate. Quando la congestione sulla rete è molto forte, i flussi in uscita dagli archi si mantengono costanti ed indipendenti dal livello di domanda. Così, differenti matrici caratterizzate da un'elevata domanda possono generare la stessa corrispondenza in termini di flusso d'arco all'interno della funzione obiettivo. Il vincolo di generazione limita l'insensibilità di alcuni termini della funzione obiettivo a tali variazioni di domanda.

Un altro aspetto da sottolineare è l'introduzione nella funzione obiettivo di pesi esterni per far lavorare in maniera corretta i singoli termini presenti in essa ed in particolare per mantenere bilanciato l'apporto di informazioni fornite dai flussi e dalle velocità.

Passando alla procedura risolutiva, nella tesi è stato adottato un metodo di ottimizzazione di tipo path search: il Simultaneous Perturbation Stochastic Approximation (SPSA, Spall 1998-1999), metodo basato su un'approssimazione del gradiente effettuata attraverso la perturbazione

simultanea di ogni variabile. Tale metodo risulta adatto a risolvere problemi a larga scala, non lineari, non analitici e caratterizzati da fonti di stocasticità, tutte caratteristiche proprie del problema di stima della domanda e dunque già di per sé si presta alla sua soluzione. Tuttavia, l'SPSA si mantiene nell'ottica di una ricerca locale e perciò la sua scelta deriva anche dalle ipotesi effettuate circa la bontà delle matrici target. Le matrici target derivano originariamente da informazioni censuarie e/o da modello, eventualmente ricostruite in statica attraverso la sola informazione dei flussi d'arco; si passa poi ad una suddivisione della matrice statica nei diversi intervalli temporali in funzione dell'andamento temporale delle misure di traffico a disposizione.

Le principali pecche di tale domanda di partenza sono da ricondurre al processo di correzione in statica che, qualora non vada a tener conto della capacità della rete, tende a sovrastimare i valori dei flussi OD ed alla suddivisione in intervalli temporali, che presuppone la conoscenza degli orari di inizio degli spostamenti. Comunque, anche in virtù del fatto che nell'usuale pratica ingegneristica la domanda di trasporto è il presupposto fondamentale per la simulazione e la seguente attività di valutazione (e quindi per qualsiasi studio di traffico), è sempre possibile ipotizzare di partire da una precedente informazione sulla domanda sufficientemente buona.

Le modifiche apportate all'algoritmo SPSA riguardano sia la fase di calcolo della stima del gradiente che la ricerca del minimo lungo la direzione di discesa e permettono di ridurre ulteriormente gli oneri computazionali della procedura e di evitare la calibrazione da caso a caso dei coefficienti presenti nella legge proposta da Spall per il calcolo del passo.

In particolare, per la stima del gradiente si è proposta la sostituzione del metodo Symmetric design (SD) con il metodo Asymmetric design (AD), il quale prevede la perturbazione del punto iniziale in una sola direzione e quindi la valutazione di una sola funzione obiettivo.

In questo modo si vanno a dimezzare il numero di valutazioni della funzione obiettivo richieste in ogni perturbazione simultanea delle componenti.

Per la ricerca della soluzione lungo la direzione di discesa sono stati proposti il metodo della sezione aurea (Golden Section Method, GSM), il quale esplora un intervallo chiuso al fine di ricercare il valore ottimo della funzione obiettivo tramite una procedura iterativa basata su una riduzione

sequenziale dell'intervallo di incertezza iniziale ed un'approssimazione polinomiale (Polynomial Interpolation – PI) che a partire da quattro punti, calcoli l'espressione del polinomio di terzo grado interpolante; da tale funzione è possibile ricavare analiticamente il relativo punto di minimo, che viene considerato come soluzione sub-ottimale ad ogni iterazione.

Il metodo risultante che adotta la combinazione dell'Asymmetric Design e dell'approssimazione polinomiale (SPSA AD-PI) si dimostra tra tutte le possibili combinazioni risolutive il miglior compromesso tra efficacia ed efficienza ed è quindi stato successivamente applicato nello studio per l'applicazione al contesto reale della città di Calgary.

La possibilità di applicare la procedura nel caso di reti di dimensioni reali è senz'altro l'aspetto più importante dello studio; difatti gli esempi applicativi presenti in letteratura fanno sempre riferimento a casi molto semplici e qualora si cerchi il passaggio verso contesti reali, ci si scontra con tempi computazionali eccessivamente elevati.

Nel presente studio, nel caso di numero massimo di variabili analizzate, la soluzione è stata trovata in massimo 15 ore (processore dual core, 3.00 GHz) e dunque in tempi appropriati ad un contesto di tipo off-line.

Ulteriori problematiche affrontate nella ricerca sono state l'importanza della tipologia, della qualità e della quantità delle misure di traffico utilizzate nella procedura di correzione, nonché la problematica dello scaricamento della rete, condizione essenziale per raggiungere un'ottimale corrispondenza in termini di flussi e velocità sulla rete, ma non ottenibile in condizioni reali.

Numerose sono ancora le prospettive di ricerca per il problema trattato, prospettive dovute all'intrinseca complessità del problema stesso.

Innanzitutto molto può essere fatto dal punto di vista dei tempi computazionali effettuando un'analisi di sensitività del parametro m che rappresenta il numero di volte in cui effettuare la stima del gradiente e che deve essere in valore inferiore al numero di variabili del problema.

Il secondo passo può essere fatto in termini di funzione obiettivo, andando ad inserire ulteriori informazioni di diversa natura (precedenti matrici censuarie, risultati indagini campionarie, informazioni da sistemi ITS) e puntando ad una normalizzazione dei singoli termini che non ricorra

all'utilizzo di pesi esterni (e che non rimanga dunque troppo ancorata alla "sensibilità" dell'analista).

Il vincolo di generazione ha dimostrato nello studio la sua effettiva validità nella ricostruzione della domanda reale; si può allora pensare di effettuare un ulteriore passo aggiungendo altre informazioni aggregate sulla domanda ed in particolare:

- in analogia con le generazioni, informazioni circa le attrazioni da modello (teoria dell'inizio dello spostamento/interviste a destinazione);
- informazioni sulla domanda complessiva a livello di sottomatrici (dato noto tramite ad esempio l'effettuazione di indagini al cordone).

In ultima considerazione, l'approccio utilizzato vede la simulazione come una "scatola nera", derivando da essa i valori di flussi e velocità necessari per la valutazione della funzione obiettivo. Il problema può essere quindi considerato come un problema bilivello in cui viene dapprima risolto il problema di assegnazione direttamente in simulazione e quindi quello di correzione della domanda; il passo successivo potrebbe essere in tal senso dato dalla soluzione del problema congiunto di correzione della domanda e dei parametri del modello che regolano il processo di assegnazione.

7. Bibliografia

- [1] Ashok, K. (1996). Estimation and prediction of time-dependent origin-destinations flows. PHD thesis. Massachusetts Institute of Technology.
- [2] Balakrishna, R. (2006). Off-line calibration of dynamic traffic assignment models. PHD thesis. Massachusetts Institute of Technology.
- [3] Bierlaire, M. (2002). The total demand scale: a new measure of quality for static and dynamic origin–destination trip tables. *Transportation Research part B*. 36:837-850.
- [4] Bierlaire, M., Crittin, F. (2004). An efficient algorithm for real-time estimation and prediction of dynamic od tables. *Operations research*. vol.52 n.1 (pp.116-127).
- [5] Box, M. J. (1965). A New Method of Constrained Optimization and a Comparison With Other Methods. *Computer Journal*, 8(1):42-52.
- [6] Brandriss, J. J. (2001). Estimation of Origin-Destination Flows for Dynamic Traffic Assignment. PHD thesis. Massachusetts Institute of Technology.
- [7] Carey, M. (1987). Optimal time varying flows on congested networks. *Operations research* 35(1), 58-69
- [8] Casas, J. (2004). Estimation of the traffic demand for real-time applications. PHD thesis. Catalunya Polytechnic University.
- [9] Cascetta, E. (1992). I modelli di assegnazione dinamica per la pianificazione e il controllo delle reti di trasporto. In “Modelli e Metodi per l’analisi ed il controllo delle reti di trasporto” a cura di E. Cascetta e G.Salerno. Franco Angeli. Milano. Italia.

- [10] Cascetta, E. (1998), Teoria e metodi dell'Ingegneria dei Sistemi di Trasporto. UTET
- [11] Cascetta, E. (2001). Transportation systems engineering: theory and methods. Kluwer Academic Publisher.
- [12] Cascetta, E., Russo, F. (1997). *Calibrating aggregate travel demand models with traffic counts: estimators and statistical performance* Transportation 24:271-293
- [13] Cascetta, E., Inaudi, D., Marquis, G. (1993). Dynamic estimators of origin-destination matrices using traffic counts. *Transportation Science* 27(4): 363-373.
- [14] Chen, H.K. and C.F. Hsueh (1998). A model and an algorithm for the dynamic user-optimal route choice problem. *Transportation Research* 32B(3), 219-234.
- [15] Cipriani, E., Florian, M., Mahut, M., Nigro, M. (2008). Investigating the efficiency of a gradient approximation approach for solution of dynamic demand estimation problem. DTA Symposium, Leuven.
- [16] Eisenman, S.M., List, G.F. (2004). Using probe data to estimate OD Matrices. Intelligent Transportation Systems Conference. Washington DC October 3-6: 291-296.
- [17] Florian, M., Mahut, M., Tremblay, N. (2006). A simulation-based dynamic traffic assignment model: Dynameq. Proceedings of the First International Symposium on Dynamic Traffic Assignment DTA2006. Institute for Transport Studies - University of Leeds.
- [18] Florian, M., Mahut, M., Tremblay, N. (2006). A simulation based dynamic traffic assignment: the model, solution algorithm and applications. Proceedings of the International Symposium of Transport Simulation ISTS06. Ecole Polytechnique Fédérale de Lausanne, Switzerland.

-
- [19] Friesz, T.L., Bernstein, D., Mehta, N.J., Tobin, R.L. and Wie, B.W. (1989). Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research* 37(6) 893-901.
- [20] Friesz, T.L., Bernstein, D., Mehta, N.J., Tobin, R.L. and and Wie, B.W. (1993). A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research* 41(1) 179-191.
- [21] Hellinga, B. R. (1994). Estimating dynamic origin-destination demands from link and probe counts. PHD thesis. Queen's University.
- [22] Hooke, R. and Jeeves, T. A. (1961). Direct Search Solution of Numerical and Statistical Problems. *Journal of the ACM*, 8:212-229.
- [23] Huyer, W. and Neumaier, A. (2004). SNOBFIT - Stable Noisy Optimization by Branch and Fit. Submitted to *ACM Transactions on Mathematical Software*.
- [24] Kattan, L., Abdulhai, B. (2006). Non iterative approach to dynamic traffic origin/destination estimation using parallel evolutionary algorithms. *Transportation Research Record: Journal of the Transportation Research Board*. vol. 1964/2006:201-210.
- [25] Kiefer, J. and Wolfowitz, J. (1952). Stochastic Estimation of the Maximum of a Regression Function. *Annals of Mathematical Statistics*.
- [26] Kleijnen, J. P. C. (1987). *Statistical Tools for Simulation Practicioners*. Marcel Dekker.
- [27] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization*, 9(1):112-147.

- [28] Lawphongpanich, S. and Hearn, D.W. (1984). Simplicial Decomposition of the Asymmetric Traffic Assignment Problem. *Transportation Research B*, 17: 123-133.
- [29] Lin, P. (2006). A robust model for estimating freeway dynamic origin-destination matrices. PHD thesis. University of Maryland.
- [30] Lin, P., Chang, G. (2007). A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix. *Transportation Research part B*. 41:554-572.
- [31] Mahut, M. (2000). Discrete flow model for dynamic network loading. Ph.D. Thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal. Published by the Center for Research on Transportation (CRT), University of Montreal.
- [32] Mahut, M., Florian, M., Tremblay, N. (2007). Comparison of assignment methods for simulation-based dynamic-equilibrium traffic assignment. Proceedings of TRISTAN VI. Phuket Island, Thailand.
- [33] Merchant, D.K. and G.L. Nemhauser (1978). A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science* 12(3), 183-199.
- [34] Merchant, D.K. and G.L. Nemhauser (1978). Optimality conditions for a dynamic traffic assignment model. *Transportation Science* 12(3), 200-207.
- [35] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M., Teller, A. H. and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**, 1087-1092.
- [36] Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *Computer Journal*, 7(4):308-313.
- [37] Newell, G.F. (1993). A simplified theory of kinematic waves in highway traffic. Part I: General Theory. *Transportation Research B*, 27B(4):281-313.

-
- [38] Nie, Y. (2006). A Variational Inequality Approach For Inferring Dynamic Origin-Destination Travel Demands. PHD thesis. University of California.
- [39] Peeta, S., Ziliaskopoulos A.K. (2001). Foundations of Dynamic Traffic Assignment: the past, the present and the future. *Networks and spatial economics*, 1:2001 233-265.
- [40] Pratelli, A. (2007). *Ingegneria dei sistemi di trasporto*. Pitagora Editrice. Bologna. Italia.
- [41] Qin, X. (2006). Traffic flow modelling with real time data for on-line network traffic estimation and prediction. PHD thesis. University of Maryland.
- [42] Ran, B. and D. Boyce (1996). *Modeling Dynamic Transportation Networks*. Berlin: Springer-Verlag.
- [43] Ran, B. and D. Boyce (1996). A link-based variational inequality formulation of ideal dynamic user-optimal route choice problem. *Transportation Research* 4C(1), 1-12.
- [44] Spall, J.C. (1998). An Overview of the Simultaneous Perturbation Method for Efficient Optimization. *Johns Hopkins APL Technical Digest*. vol. 19, Number 4 (pp. 482-492).
- [45] Spall, J.C. (1998). Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on aerospace and electronic systems*. vol. 34 n.3 July 1998.
- [46] Van der Zijpp, N.J. (1996). *Dynamic Origin-Destination Matrix Estimation on Motorway Networks*. PhD thesis. Delft University of Technology. Department of Civil Engineering.
- [47] Wie, B.W. (1990). Dynamic Analysis of user optimized network flows with elastic travel demand. Presented at the 70th TRB Annual Meeting, Washington, D.C.

- [48] Zhou, X., Mahmassani, H.S. (2006). Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Transactions on intelligent transportation systems*. vol.7 n.1:105-114.

- [49] Zhou, X., Mahmassani, H.S. (2007). A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework. *Transportation Research part B*. 41:823-840.

Ringraziamenti

Quando si giunge alla chiusura di un “piccolo capitolo” della vita, ci si rende conto che sono sempre molte le persone che hanno contribuito alla sua realizzazione.

In queste poche righe vorrei poterle ringraziare tutte:

grazie al Prof. Stefano Gori ed all’Ing. Ernesto Cipriani che mi hanno seguita con costanza e tanta (e sottolineo tanta) pazienza in questo difficile percorso;

grazie all’Ing. Marco Petrelli per avermi fatto capire ogni giorno quanto sia importante avvicinarsi alla ricerca con grande modestia;

grazie al prof. Michael Florian per avermi ospitata e seguita presso il “Centre de recherche sur les transports” dell’Università di Montreal. Grazie per avermi dato questa opportunità, è stato per me un periodo di grande crescita;

grazie a Michael Mahut e Nicolas Tremblay della INRO Consultants per il loro sempre pronto supporto;

grazie al prof. Giulio Erberto Cantarella per tutti i suoi consigli e suggerimenti e per aver sempre sostenuto quanti sono solo ai loro primi passi nel mondo della ricerca;

grazie a tutti coloro che hanno reso “vivo” il Laboratorio Trasporti del DSIC Roma Tre, mia seconda casa in questi quattro anni, a chi vi è solo passato e a chi vi si è fermato;

grazie a tutte le persone che ho potuto conoscere nei miei viaggi di studio in Italia ed all’Estero e che mi hanno resa partecipe di un mondo affascinante come quello della ricerca;

grazie a tutti i miei studenti, che con le loro domande, le loro osservazioni, le loro curiosità mi hanno sempre spinto ad approfondire ed a migliorare le mie conoscenze;

grazie alla mia famiglia che mi ha sempre supportata nei momenti felici ed incoraggiata nei momenti difficili;

grazie a Marco, il mio passato, il mio presente, il mio futuro.