

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

Scuola di Scienze
Corso di Laurea Magistrale in Fisica

**RICOSTRUZIONE DI FLUSSI VEICOLARI
SU SCALA REGIONALE:
PROGETTAZIONE E IMPLEMENTAZIONE
DI UN MODELLO**

Relatore:
Prof. Sandro Rambaldi

Presentata da:
Marco Di Cristina

Sessione II
Anno Accademico 2012/2013

Abstract

Si mostrano i passi seguiti nello sviluppo di un modello matematico e simulativo che mira a ricostruire il campo di flusso veicolare della rete stradale della regione Emilia Romagna. Si descrivono le fonti di dati utilizzate per la ricostruzione: dispositivi GPS a bassa capillarità e spire induttive su strada; si discute delle caratteristiche dei dati e dei problemi intrinseci e contingenti da cui sono affetti, nonché delle metodologie adottate nel tentativo di risolvere questi problemi. Si illustra l'idea centrale alla base delle ricostruzioni, e si descrive un criterio di validazione del modello. Si mostrano i risultati dei vari algoritmi ideati, ognuno affiancato dal parametro di validità, evidenziando quale metodo produca risultati migliori. Si conclude discutendo delle potenzialità del modello in termini di applicabilità in altre regioni o zone geografiche.

Indice

1	L'oggetto di studio	4
1.1	La problematica di fondo	5
1.2	Il modello	9
2	Descrizione dei dati	11
2.1	Selezione dei dati	11
2.2	Caratteristiche giornaliere dei flussi di traffico	13
2.3	Le fonti di dati	15
2.3.1	Spire	15
2.3.2	Dati GPS	22
2.4	La penetrazione dei dati GPS	29
2.4.1	Calcolo in assenza di dati da spire	30
2.4.2	Calcolo con dati da spire	34
2.4.3	Conclusioni	35
3	Algoritmi di propagazione	37
3.1	Introduzione	37
3.2	Trattazione teorica	38
3.3	Modelli	39
3.3.1	Calcolo dei pesi	41
3.3.2	Propagazione dei flussi	43
3.3.3	Validazione del modello	44
3.4	Rappresentazione schematica del modello	45

4	Analisi dati	49
4.1	Correlazioni fra spire	49
4.2	Distinzione in fascia mattutina e pomeridiana	50
4.3	Parametri caratteristici delle spire	50
5	Risultati	52
5.1	Pesi con tempi di viaggio, propagazione con rapporti di flusso .	54
5.2	Pesi con tempi di viaggio, propagazione senza rapporti di flusso	58
5.3	Pesi con probabilità di svolta, propagazione con rapporti di flusso	60
5.4	Pesi con probabilità di svolta, propagazione senza rapporti di flusso	62
5.5	Sintesi dei risultati	63
5.6	Calcolo del campo di flusso nella regione	66
5.6.1	Ore 01:00	67
5.6.2	Ore 04:00	68
5.6.3	Ore 06:00	69
5.6.4	Ore 08:00	70
5.6.5	Ore 10:00	71
5.6.6	Ore 18:00	72
5.6.7	Ore 20:00	73
5.6.8	Ore 23:00	74
6	Conclusioni	75
	Bibliografia	78

Premessa

È doveroso, prima di iniziare la trattazione del lavoro svolto, parlare della divisione dei compiti e delle responsabilità nell'insieme dei vari argomenti studiati e trattati.

Il progetto è stato ideato e guidato dal Gruppo di Fisica dei Sistemi Complessi dell'Università di Bologna, e più specificamente dal Laboratorio di Fisica della Città. Lo staff del laboratorio, di cui è parte il Relatore di questa tesi, ha costantemente diretto e osservato l'evoluzione del lavoro.

L'autore di questa tesi è solo uno dei due laureandi che ha lavorato al progetto sotto la supervisione dello staff, e non è l'autore dell'intera realizzazione del modello trattato in questa tesi. I due laureandi hanno lavorato secondo una divisione dei compiti, e il progetto è frutto della sintesi di entrambi gli operati.

Nella redazione di questa tesi, si è scelto di dare un'idea completa del progetto, e tuttavia alcuni capitoli o paragrafi di questa trattazione appariranno molto sintetici; ciò è dovuto al fatto che gli argomenti di quella sezione sono stati studiati dall'altro laureando. In questi casi sarà comunque fatto presente di fare riferimento alla tesi redatta da lui.

In particolare, l'autore di questa tesi si è occupato, più dell'altro laureando, della raccolta, lettura e pre-elaborazione dei dati, nonché dello sviluppo e implementazione tecnica degli algoritmi;

Il capitolo che parla dei risultati sarà una sintesi delle conoscenze acquisite da entrambi, ed essendo opera comune apparirà in entrambe le tesi.

Introduzione

L'obiettivo principale di questo lavoro di tesi è lo sviluppo di un modello in grado di ricostruire il campo di flussi su un'ampia rete stradale a partire da misure sparse.

Il progetto si inquadra pertanto nell'ambito di quel settore multidisciplinare di studio che si occupa della mobilità di veicoli su reti stradali.

È noto che al costante sviluppo urbano delle zone metropolitane in tutti i paesi industrializzati non è sufficiente una risposta in termini di mera espansione delle infrastrutture, tra l'altro via non sempre percorribile in zone con interesse storico artistico o architettonico.

Indispensabile è un processo di miglioramento, ottimizzazione e modernizzazione delle infrastrutture di trasporto già presenti, processo che nella pratica si traduce in una modifica mirata delle vie di comunicazione esistenti, o di gestione razionalizzata dei flussi di traffico veicolare, di modifica della struttura viaria della rete, o di riorganizzazione geografica, quando possibile, dei luoghi che costituiscono bacini di attrazione del traffico.

Quali che siano le soluzioni che si intende intraprendere, il processo non può prescindere da un estensiva e costante osservazione e analisi del sistema viario dal punto di vista di molte proprietà legate ad esso.

Molte discipline scientifiche hanno sviluppato un interesse, nell'ultimo mezzo secolo, in questo settore di studio, non da ultima la fisica in varie sue branche: fisica matematica, fisica dei sistemi dinamici, dei sistemi complessi.

Nell'analisi dei sistemi dinamici di questo tipo possono essere adottati,

traendo spunto dalla fluidodinamica[1], due approcci osservativi: quello lagrangiano e quello euleriano.

Nel quadro lagrangiano si concentra l'osservazione sull'individuo, seguendo la dinamica all'interno del sistema; in quello euleriano si osservano invece le grandezze in un punto del sistema studiandone le variazioni allo scorrere del tempo.

La simulazione e la modellizzazione sono elementi fondamentali nello studio dei sistemi di traffico, e gli approcci utilizzati dalla fisica applicata alle simulazioni di reti urbane si diramano in due categorie, che costituiscono una gerarchia dal punto di vista della scala osservativa dei fenomeni[2].

L'approccio macroscopico lascia da parte le proprietà microscopiche come velocità dei singoli individui, mantenendo un punto di vista globale del problema. Nel quadro lagrangiano ciò si tradurrebbe nello studio di flussi tra grandi gruppi di individui, ad esempio città o addirittura regioni. Dal punto di vista euleriano si avrebbe uno studio dell'evoluzione temporale delle grandezze di questi gruppi.

Quello microscopico tratta in dettaglio le interazioni e le dinamiche dei singoli componenti, seguendo ogni individuo (che in questo caso è un veicolo) nello spazio e nel tempo.

Questa tesi descrive un approccio euleriano in cui la grandezza fondamentale su cui si concentra il modello è il flusso di veicoli: il numero di veicoli che transita nella rete per unità di tempo. Va fatto notare, tuttavia, che il modello proposto fa largo uso di dati di natura prettamente lagrangiana raccolti in precedenza; questi dati forniscono informazioni particolarmente utili per la costruzione del modello.

È importante anche considerare che nel sistema qui studiato il più piccolo oggetto osservabile non è il singolo punto, ma la singola strada. Questo è conseguenza del fatto che si è scelto di sviluppare un modello che tralasciasse informazioni dinamiche all'interno delle strade.

Capitolo 1

L'oggetto di studio

L'obiettivo principale del lavoro è la ricostruzione dell'intero campo di flusso veicolare per la regione Emilia Romagna a partire da misure locali e sparse raccolte con alta frequenza.

Una descrizione sufficientemente completa per questo particolare studio della rete stradale può essere fornita utilizzando la teoria dei grafi[3], trattando semplicemente ogni strada come un link ed ogni incrocio come un nodo; questa semplificazione può essere applicata perché l'interesse dello studio non ricade su una risoluzione della dinamica individuale del singolo veicolo a livello microscopico, ma sulla dinamica integrata su un'intera strada; le dinamiche interne di una strada vengono riassunte da un set di valori che ne descrivono l'andamento medio.

In quest'ottica il network che rappresenta la rete urbana è un grafo diretto in cui le strade, rappresentate da collegamenti, link o archi, sono direzionate e gli incroci tra una strada e l'altra sono i nodi.

Molte caratteristiche delle reti cittadine, quali ad esempio l'agibilità dei possibili percorsi, la rapidità di percorrenza o la capacità in termini di densità di mezzi, possono essere studiate mediante l'analisi delle proprietà topologiche del network. Queste proprietà parametrizzano il sistema con un basso numero di grandezze, rendendolo un più semplice oggetto di studio.

Un approccio Euleriano al problema richiederebbe lo studio di varie quantità, principalmente il campo di flusso, di velocità e di densità. La sua descrizione matematica utilizza sistemi di equazioni differenziali alle derivate parziali e fornisce leggi di conservazione o di bilancio. In questo studio, tuttavia, ci si è concentrati sul campo di flusso.

La conoscenza del flusso stradale per l'intera rete è, evidentemente, un'informazione molto preziosa per la gestione del traffico poiché fornisce indicazioni sulle zone in cui gli spostamenti avvengono ed evolvono in maniera fluida e quelle dove invece il traffico è soggetto a congestioni; di conseguenza facilita un'analisi mirata al miglioramento del traffico veicolare.

Misure dirette e onnicomprensive sulla rete sono però, per ragioni principalmente economiche, difficili da realizzare e di conseguenza si effettuano misure *sparse*, solo in opportuni punti del network; il problema di estendere l'informazione spaziale sui flussi all'intera rete a partire da un numero limitato di misure non è univocamente determinato e richiede ulteriori informazioni.

Nello studio presente, i dati a disposizione provengono da apparati di misura fissi in grado di rilevare i veicoli che transitano sulla strada. Questi dispositivi, di forma solenoidale e denominati *spire*, sono posizionati sotto il manto stradale. Queste spire hanno la possibilità, oltre a conteggiare i veicoli, di valutarne la velocità e di classificarli, basandosi sul tempo di transito ad altri fattori, in diverse categorie veicolari.

Gli apparati sono stati installati, dall'ente regionale per la mobilità, sulle strade principali della rete regionale; il loro funzionamento verrà spiegato più in dettaglio nel Capitolo 2.

1.1 La problematica di fondo

Il problema è, dal punto di vista fisico/matematico, quello della ricostruzione temporale di un campo scalare su un dominio che ha una topologia a network, conoscendo il valore del campo solo in alcuni punti.

Appare subito evidente che il problema è molto sottodeterminato, in quanto il network regionale si compone di almeno 100 mila strade, e sono a disposizione poco più che 270 postazioni, che forniscono, considerando i due versi di transito, 540 dati, di cui almeno 40 con differenti patologie che li rendono di difficile utilizzo. Il campo, in queste condizioni, non è risolvibile univocamente.

Per estendere l'informazione alle zone circostanti alle misure puntuali sono state incluse tutte le informazioni dinamiche di cui si ha conoscenza ed è stata sviluppata una metodologia per misurare l'errore delle ricostruzioni.

Una proprietà del sistema che semplifica la trattazione è la topologia a network, che ne riduce la dimensionalità, rendendolo quasi unidimensionale.

Nell'immagine 1.1, che mostra l'ingrandimento di un settore del network, in rosso sono evidenziati alcuni nodi, mentre in grigio i collegamenti.

Figura 1.1: Un settore di network stradale



Inoltre, il campo di flusso soddisfa una legge di continuità in virtù del semplice fatto che gli autoveicoli si conservano: si assume che gli individui non spariscano né nascano negli incroci, ma possano sparire o nascere lungo un tratto di strada. La presenza di soste e partenze complica il problema introducendo *pozzi* e *sorgenti* nel campo. Le posizioni di questi pozzi e sorgenti sono ignote e sono una delle principali fonti di errore del modello.

In breve, il problema trattato è quello del trasporto e conservazione di una grandezza in un sistema quasi unidimensionale con sorgenti e pozzi diffusi e ignoti.

I dati a disposizione sono un set limitato di misure di questa grandezza ed una conoscenza storica.

I dati provenienti dalle spire possono essere considerati privi d'incertezza, poiché tutte le altre sorgenti d'errore del modello sono molto maggiori.

La conoscenza storica del sistema fa da guida nell'estensione del campo. Tale conoscenza deriva da dati di posizione GPS raccolti da sensori di cui si sono dotati per ragioni assicurative, in maniera volontaria, una percentuale dei veicoli circolanti su tutta la rete nazionale; la percentuale di autoveicoli dotati di GPS, vale a dire la *penetrazione* di questo campione, è approssimativamente il 2% di tutti i veicoli immatricolati.

Se si avessero tutti i dati da GPS si avrebbe di conseguenza la conoscenza dello stato di ogni veicolo in tempo reale e i flussi sarebbero perfettamente noti.

Si è verificato che nei punti in cui esistono le misure da spira la ricostruzione dei flussi a partire dai dati GPS non coincide con i flussi calcolati dalle spire. Questa differenza dipende dalle fluttuazioni spaziali e temporali della percentuale di autoveicoli con GPS rispetto ai veicoli totali. Queste fluttuazioni sono molto alte nelle zone a basso traffico.

Questi dati serviranno per lo sviluppo del modello di estensione dei flussi, che farà uso, come dettagliato nel Capitolo 3, dei rapporti fra i flussi di ogni coppia di strade, che quindi non può prescindere da una conoscenza almeno parziale di tutti i flussi.

Ovviamente se tutte le auto avessero un rivelatore GPS la conoscenza del campo di flusso sarebbe perfetta.

Se, idealmente, la distribuzione dei GPS fosse omogenea su tutta la regione, vale a dire se il valore di penetrazione fosse costante ed unico per tutte le strade, questo campione di veicoli sarebbe rappresentativo e si stimerebbero i flussi reali con un fattore moltiplicativo.

Un'altra informazione di cui si deve tener conto è la conoscenza storica del rapporto fra i flussi delle varie strade, che deve fare da termine di paragone

per le ricostruzioni successive. Se fossero noti tutti questi rapporti ed essi si mantenessero costanti nel tempo, basterebbe la conoscenza del flusso in una strada per mappare l'informazione sull'intero sistema in esame.

All'atto pratico si è verificato che la penetrazione di questi dati GPS è continua e si può considerare omogenea in intorni spaziali dell'ordine di 10 km e questo è plausibile in quanto la maggior parte degli individui in media si muovono localmente in un intorno della residenza o del luogo di lavoro. Inoltre il numero di individui possessori di GPS di una zona varia lentamente nel tempo. Si è inoltre verificato che la penetrazione non varia in maniera significativa tra le ore di punta della giornata (mattina 7:00 - 9:00 e pomeriggio 17:00 - 19:00) quindi si sono utilizzati dati medi mensili per stimare come la penetrazione vari spazialmente.

Un'altra quantità che risulta importante per l'analisi condotta è il campo di velocità. Questa quantità è molto più facile da stimare dei flussi a partire dai dati GPS. Ad esempio, quale che sia la percentuale di individui con GPS su un particolare tratto, essi si comporteranno come tutti gli altri individui sulla strada: se è molto trafficata, terranno una bassa velocità media, se è scorrevole la manterranno più alta. Per questa grandezza i dati sono corretti indipendentemente dalla penetrazione del nostro campione, che in questo caso svolge il ruolo di stabilizzare il valor medio, ossia quando la penetrazione è alta si hanno molti più dati e quindi il valor medio fluttua meno. La deviazione standard nel campo di velocità è dell'ordine di 30 Km/h ed è dovuta al differente comportamento degli individui.

Per riassumere, si hanno due fonti di dati per il sistema, ed entrambe le classi di dati sono utilizzate nella costruzione del modello:

- quelli che provengono dalle spire sono dati sparsi, limitati in numero ma, a meno di guasti tecnici, accurati;
- quelli provenienti dai GPS sono dati numerosi e diffusi spazialmente, ma di difficile utilizzo soprattutto per la loro disomogeneità spaziale.

L'idea alla base di questo studio è quindi quella di sintetizzare la cono-

scenza completa del sistema, unendo una conoscenza scarsa ma diffusa ad una accurata ma puntuale.

1.2 Il modello

Il modello sviluppato fa uso di tutti i dati descritti sopra: viene sviluppata una dinamica spaziale e temporale dei flussi, propagando l'informazione iniziale sulla rete e ricevendo a intervalli regolari i dati dalle spire. Alla fine del processo si valuta la bontà della ricostruzione.

Come fase preliminare si sono calcolati a partire dai dati GPS i flussi nei punti esatti dove erano presenti le spire e questo ha permesso di valutare la penetrazione su queste strade (che sono le strade principali dell'Emilia Romagna) ed ha evidenziato come l'assunzione, che era stata fatta in precedenza, di penetrazione essenzialmente costante fosse una grande sorgente di errore. Si veda la Fig. 1.2.

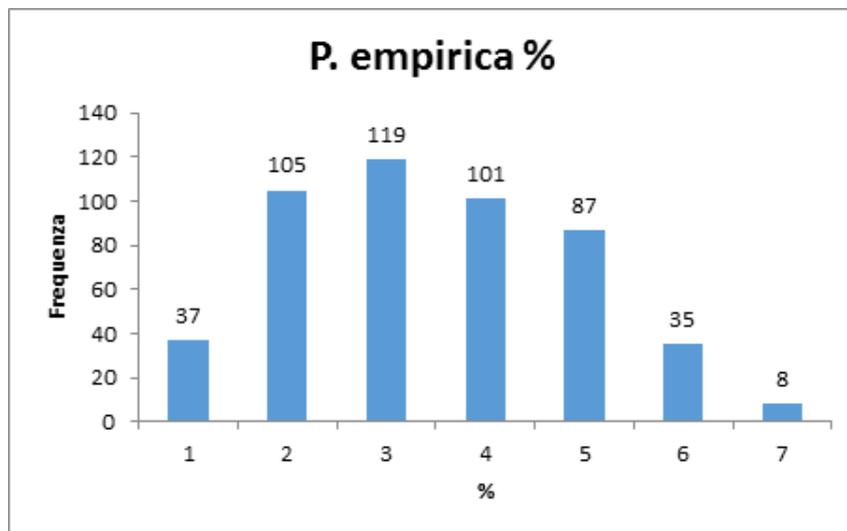


Figura 1.2: Distribuzione dei valori di Penetrazione empirica. Per i dettagli sul calcolo si veda il Capitolo 2.

Ciò ha anche evidenziato che una precedente stima della penetrazione locale a partire dalla attribuzione di un comune di residenza ad ogni singolo

veicolo dotato di GPS e dal confronto con il numero di auto immatricolate nel medesimo comune risultava molto più precisa.

Di conseguenza si è utilizzata questa stima della penetrazione per valutare il campo di flusso a partire dai dati GPS. In particolare si sono calcolati due campi medi: uno per l'ora di punta mattutina e l'altro per quella pomeridiana.

Da qui si è sviluppato il modello matematico in grado di propagare l'informazione da ogni strada alle successive.

La base del modello di calcolo è la seguente e si divide in due passi:

- nella prima fase per ogni strada viene calcolata la dipendenza dalle strade vicine. Sono stati sviluppati vari algoritmi che svolgono questo compito in maniera diversa e sono stati confrontati;
- nella seconda è stata propagata l'informazione del flusso dalle strade con le spire verso tutte le altre utilizzando i risultati della prima fase.

Naturalmente il modello progettato non è perfetto, e non produce risultati totalmente accurati. Non ci si può esimere dall'introdurre nel modello una fase di validazione, cioè di verifica della bontà dei risultati.

In questo lavoro il processo di validazione si sviluppa nei seguenti passi:

- ricostruzione dei flussi sulle strade dotate di spira come risultato della propagazione del campo globale dalle spire vicine;
- verifica dell'accuratezza della previsione mediante il confronto dei flussi ricostruiti con quelli misurati dalla spira.

Il modello, qui solo brevemente descritto, viene illustrato più in dettaglio nel Capitolo 3.

Capitolo 2

Descrizione dei dati

2.1 Selezione dei dati

Come introdotto in precedenza, questo lavoro di tesi si sviluppa a partire da un limitato data set iniziale, e procede nel tentativo di ricostruzione dei flussi sull'intera rete regionale tramite un modello computazionale che restituisce risultati in tempo reale.

Le tipologie di dati a disposizione sono due, di natura diversa:

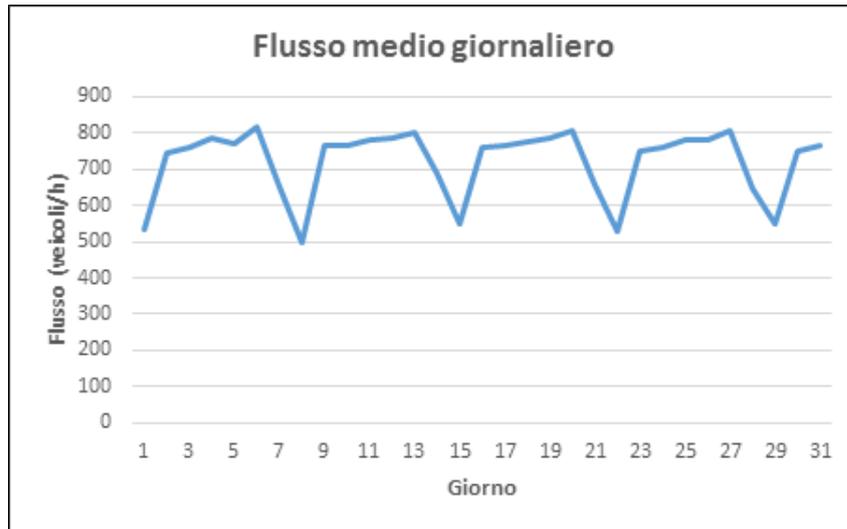
- dati da GPS, appartenenti ad una certa percentuale di tutte le auto immatricolate nella regione, relativi al mese di maggio 2011;
- dati da postazioni fisse su strada, di tutto il 2011.

Lo scopo della selezione ed elaborazione di questi dati è quello di ottenere un insieme coerente di dati, al fine di creare un modello di calcolo e simulazione efficace. Quindi si è scelto di selezionare, per entrambi i set, dati di periodi corrispondenti.

Inizialmente si è fatta una selezione per tutti i giorni del mese di maggio 2011, ma poiché lo scopo era un'analisi dei dati che evidenziasse le caratteristiche che meglio permettessero di risolvere il sistema, è preferibile che i dati non presentino troppe oscillazioni durante il periodo considerato. Per questo si è

scelto di dividere il mese in giornate feriali e giornate festive, ipotizzando che il comportamento degli automobilisti potesse variare fra queste due tipologie di giornata[4].

Figura 2.1: Flusso medio giornaliero.



Per esemplificare questo concetto, in Fig. 2.1 si mostra l'andamento del valor medio di flusso giornaliero, per tutti giorni del mese di maggio 2011, di una scelta arbitraria di spira.

Poiché l'interesse di un'analisi di questo tipo era la selezione di una buona statistica dei dati, si è scelto di utilizzare le giornate feriali, poiché maggiori in numero e con meno variazioni dovute a motivi indipendenti dalla natura del sistema.

Si è visto, infatti, che durante i fine settimana o i giorni feriali i flussi di veicoli sulla rete regionale sono molto più variabili e imprevedibili, poiché gli spostamenti sono governati da fattori esterni alla routine quotidiana. Per renderne conto correttamente, si sarebbe dovuta tenere in considerazione la presenza di fenomeni atmosferici, festività, ponti ed altre variabili che avrebbero complicato notevolmente il modello.

2.2 Caratteristiche giornaliere dei flussi di traffico

I dati analizzati per l'implementazione del modello sono, come si vedrà in dettaglio più avanti, andamenti di flusso giornaliero, relativi ad una particolare posizione geografica della regione, discretizzati nel tempo.

Nella Fig. 2.2 è mostrato l'andamento giornaliero del flusso orario ottenuto mediando gli andamenti delle prime 100 postazioni più popolate del giorno 4 maggio 2011. Questo grafico dà un'idea realistica di come mediamente il traffico si evolva su ogni strada durante la giornata.

Figura 2.2: Flusso tipico.



Osservando questo e altri grafici analoghi si possono trovare diverse caratteristiche dell'andamento medio. In particolare, il flusso giornaliero appare articolato in quattro fasi:

- la prima è una sorta di plateau nelle ore notturne, in cui il flusso rimane basso;
- poi è presente un picco mattutino che inizia ad alzarsi dalle 06:00 fino ad arrivare al suo massimo attorno alle 08:00, poi calare fino alle 09:00;

- poi una zona più o meno contenuta e più o meno costante durante la giornata, con due picchi secondari attorno all'ora di pranzo, fino ad arrivare alle 16:00, dove il flusso aumenta nuovamente crescendo fino alle 18:00 ove raggiunge il valore massimo;
- infine la discesa fino a verso le 20:00.

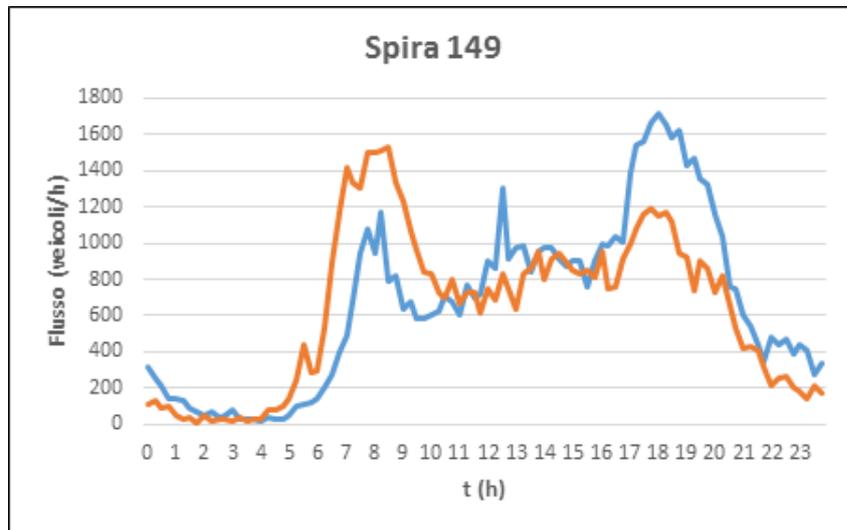
Queste informazioni sono di relativamente facile interpretazione. Nelle ore mattutine la tendenza degli automobilisti è quella di recarsi al posto di lavoro, dinamica che si mantiene fin verso le 09:00, quando si osserva che il flusso diminuisce e rimane stabile fino a metà giornata. Qui è possibile osservare i due picchi secondari, verso le 12:00 e 14:30, caratteristici della tendenza di una parte degli individui a spostarsi in auto per la pausa pranzo, o che possono suggerire l'inizio o la fine di giornate lavorative part-time. Fino alle 16:00 il flusso si mantiene all'incirca costante, fino al picco delle 18:00 che indica la fine della giornata lavorativa. Da qui il flusso cala costantemente fino a circa l'01:00-02:00 di notte, dove inizia il plateau notturno. È interessante porre l'attenzione sulla differenza dei due picchi principali in termini di larghezza: la differenza è indicativa della tendenza degli individui a compiere attività esterne al lavoro in orari pomeridiani/serali; la mattina questo comportamento non si osserva, poiché la tendenza è quella, appena usciti di casa, di recarsi direttamente al posto di lavoro.

Sebbene il grafico mostrato in precedenza mostri l'andamento medio, non si può assumere che tutte le strade rispettino questa tendenza. Si è verificato che ci sono strade più trafficate la mattina rispetto alla fascia serale o viceversa, mostrando, in termini di grafici, altezze diverse fra i picchi.

Ad esempio di ciò, si osservi nella Fig. 2.3 il flusso giornaliero di una postazione, in cui appaiono entrambi i sensi di marcia della strada.

Si osserva che nei due versi le ampiezze dei due picchi siano speculari, indicando, in questo caso, un'affluenza maggiore al mattino per una corsia e maggiore al pomeriggio per l'altra.

Figura 2.3: Una spira che presenta asimmetria fra le due direzioni di percorrenza.



Si sono verificate anche altre eccezioni, quali ad esempio strade che non presentano affatto uno dei due picchi.

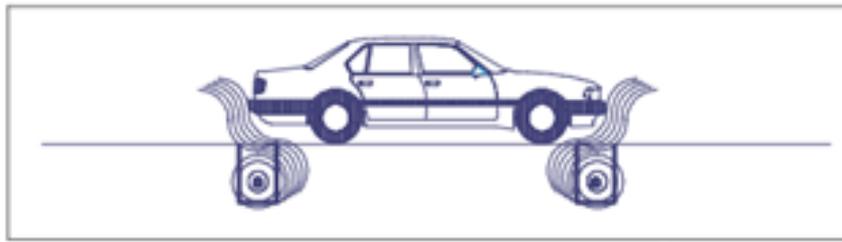
L'andamento medio non è sempre ed ovunque rispettato, ma studiando un grande numero di andamenti si osserva che il sistema regionale si comporta in maniera molto isocrona, mostrando molte ricorrenze nelle variazioni dei flussi.

2.3 Le fonti di dati

Come già affermato, le due fonti di dati utilizzate in questo progetto sono postazioni fisse e sparse, che d'ora in poi nomineremo spire, e GPS più o meno diffusi su tutta la regione. In questa sezione si parlerà più in dettaglio di queste due fonti di dati.

2.3.1 Spire

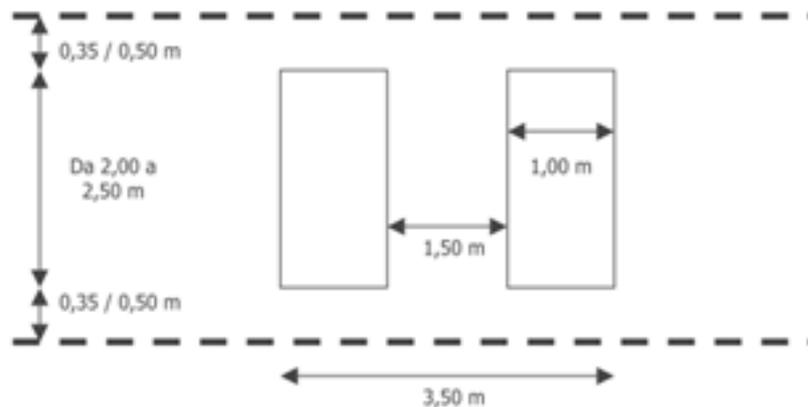
Con il termine spira si intende una postazione^[5] composta di due spire induttive per senso di marcia, posizionate al di sotto del manto stradale come schematizzato in Fig. 2.4.

Figura 2.4: Schema della rivelazione di un veicolo.

Le spire permettono di misurare, al passaggio del veicolo sopra di esse, diverse grandezze: classe veicolare, velocità, lunghezza veicolo, distanza dal veicolo precedente, tempo di percorrenza, time gap, direzione di marcia.

Installazione

Per procedere all'installazione è necessario effettuare due coppie di incisioni nel manto stradale, di profondità 50-70 mm e di larghezza di 8-10 mm, componendo rettangoli di circa 2-2.5 m per 1 m distanti circa 1.5 m, come in Fig. 2.5.

Figura 2.5: Incisioni del manto stradale per l'installazione.

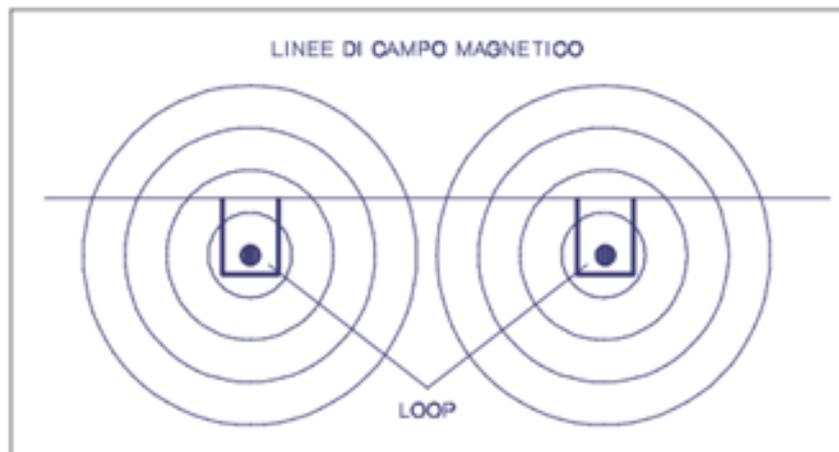
Quindi vengono pulite le scanalature e vi vengono depositati i cavi, arrotolati 50-100 volte, in cui passerà la corrente. Occorre prestare particolare attenzione agli angoli dei loop, che devono essere smussati per non rovinare il cavo, poiché sotto tensione potrebbe rovinarsi. Posizionati come in figura si

procede alla cementificazione. Entrambi i lembi di cavo che escono dai due lati del loop vengono collegati ad un rivelatore posizionato a bordo stradale.

Funzionamento

I due set di cavi arrotolati nel terreno sono percorsi da correnti alternate le quali inducono un campo magnetico, anch'esso di conseguenza alternato, come si nota in Fig. 2.6. Quando un veicolo si trova in prossimità del campo magnetico, interagisce con esso, e vengono indotte all'interno del veicolo stesso delle correnti le quali a loro volta generano un campo magnetico che si va a contrapporre a quello preesistente. In funzione di questo fenomeno l'induttanza del campo magnetico indotto diminuisce e di conseguenza la frequenza di oscillazione aumenta.

Figura 2.6: Linee di campo magnetico generato dalle spire.



Uno strumento rileva il numero di oscillazioni in un certo intervallo di tempo ed è in grado di verificare se un veicolo è passato al di sopra delle spire. È inoltre interessante notare che lo strumento non registra le variazioni lente delle oscillazioni di campo, dovute ad esempio a cambiamenti di temperatura, e quindi è in grado di distinguere il passaggio di veicoli da eventi di altra origine.

Le spire della regione Emilia Romagna

Nella regione Emilia Romagna sono state inserite 278 postazioni funzionanti seguendo le metodologie esposte precedentemente. È importante notare che per spira in questo lavoro si intenderà la postazione che fornisce il set di dati relativo ad una sola direzione di marcia della strada; a causa di questo ogni strada dotata di spira avrà associate, nel modello, due spire.

La scelta del posizionamento delle spire è stata effettuata dalla Regione Emilia Romagna, dalle Provincie e dall'ANAS. Le postazioni sono state posizionate nelle strade più importanti della regione, escludendo i tratti autostradali; la loro distribuzione è quindi disomogenea nello spazio, poiché la logica alla base della scelta era l'analisi delle zone più critiche della regione, e non necessariamente di ottenere dati equispaziati. Per questo molte delle spire sono su strade importanti in termini di traffico.

I risultati delle misure di queste postazioni sono conteggio dei veicoli transitati sul tratto di strada ogni 15 minuti.

I dati forniscono informazioni sul diverso tipo di autoveicoli, ma l'interesse del modello si è concentrato sulle automobili monovolume, che comunque costituiscono la maggior parte dei transiti. Il motivo di questa scelta è dovuto al fatto che i dati di GPS a disposizione comprendono principalmente le autovetture e non le altre diverse categorie, quali autobus, camion o ciclomotori.

I dati di spira appaiono, nei file messi a disposizione dalla Regione, nella seguente forma testuale:

spira; data e ora; ora; corsia; flussi nelle categorie veicolari
319;04/05/2011 18.15.00;18,25;1;0;0;2;82;0;7;1;0;2;0;0

Insieme ai dati di flusso, per ogni spira sono state fornite informazioni sul loro posizionamento geografico e stradale, nonché sulla corrispondenza dei due versi di percorrenza con direzioni geografiche assolute.

È evidente che per utilizzare i dati è stato necessario associarvi una posizione anche nella cartografia del programma di simulazione, e ciò è stato possibile

utilizzando le coordinate di localizzazione delle spire e osservando quale fra le strade più vicine a quel punto avesse un flusso non trascurabile.

Come già precisato, infatti, le spire sono posizionate su strade importanti della rete regionale: non avrebbe avuto senso assegnare una spira ad una strada secondaria, nonostante la sua vicinanza al punto di localizzazione della spira, quando altrettanto vicino o poco meno si trovava una strada importante.

Una verifica della corretta assegnazione è stata comunque possibile confrontando le informazioni testuali sulla denominazione delle strade con spire con le informazioni già contenute nella cartografia.

Per quanto riguarda le direzioni di percorrenza, invece, a causa della struttura della cartografia interna e dell'arbitrarietà dei versi di percorrenza delle spire, è stato necessario assegnare manualmente ad ogni verso di percorrenza di ogni singola spira la corretta poligonale, basandosi sulle informazioni di direzione fornite testualmente.

Elaborazione dati

Prima di poter utilizzare i dati dalle postazioni fisse è stato necessario studiarli e pulirli da eventuali errori e si è osservato che erano affetti da diversi problemi, quali principalmente:

- dati erroneamente nulli, con zeri al posto dei valori di flusso corretti, oppure campi dati completamente vuoti;
- mancanza di interi orari dal file;
- spire non funzionanti per una gran parte del periodo preso in esame;
- scambio degli identificativi tra due spire.

Per risolvere i problemi del primo tipo, dopo essere stati identificati, si è scelto di interpolare i valori in modo da rendere la funzione che rappresenta il flusso più continua possibile; quando invece, per la seconda classe di spire problematiche, il numero di dati mancanti risultava troppo grande da rendere

impossibile una interpolazione si è scelto di escludere l'intero set di dati di quella spira.

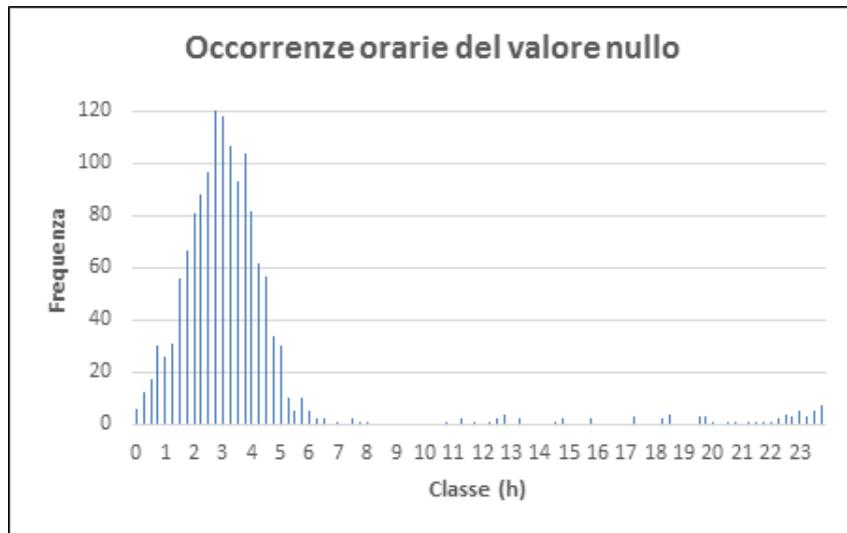
Anche il terzo problema, come presumibile, ha portato alla diretta eliminazione delle spire dal sistema, perché sarebbe stato impossibile estrarne informazione utile, se non ricorrendo ad artifici matematici che l'avrebbero con grande probabilità falsata.

L'ultimo problema è stato risolto scambiando l'identificativo delle spire.

È inoltre opportuno sottolineare che una particolare spira ha un problema di natura differente: per la sua posizione in curva, una corsia valuta un grande numero di transiti, mentre quella in verso opposto ne conteggia assai pochi; questo è molto probabilmente attribuibile al fatto che le auto, percorrendo la strada ad alta velocità nella corsia esterna, tendano a tagliare la curva non permettendo alla spira corretta di registrare il passaggio di autoveicoli. Sembra improbabile che i dati siano attribuiti alla spira posizionata nella direzione opposta, poiché gli apparati permettono di misurare anche la velocità di percorrenza degli autoveicoli, ma non conoscendo a fondo le modalità di reazione rispetto a queste percorrenze e per non rischiare di introdurre dati errati si è scelto di escludere l'intera coppia di spire.

Successivamente, dopo aver selezionato le spire che non possedevano comportamenti anomali, è stato scelto di interpolare le spire con gli zeri erronei; ciò è stato effettuato suddividendo l'intera giornata in giorno e notte, partendo dal presupposto che la presenza di un flusso pari a 0 nella fascia notturna non rappresenta un errore, mentre è molto probabilmente causato da un malfunzionamento dello strumento quando questa occorrenza avviene in pieno giorno. Nella Fig. 2.7 è mostrato l'istogramma degli zeri. Si osserva che nella fascia di ore dalle 22:00 alle 7:00 gli zeri sono molto diffusi; questa fascia è stata effettivamente scelta come quella notturna.

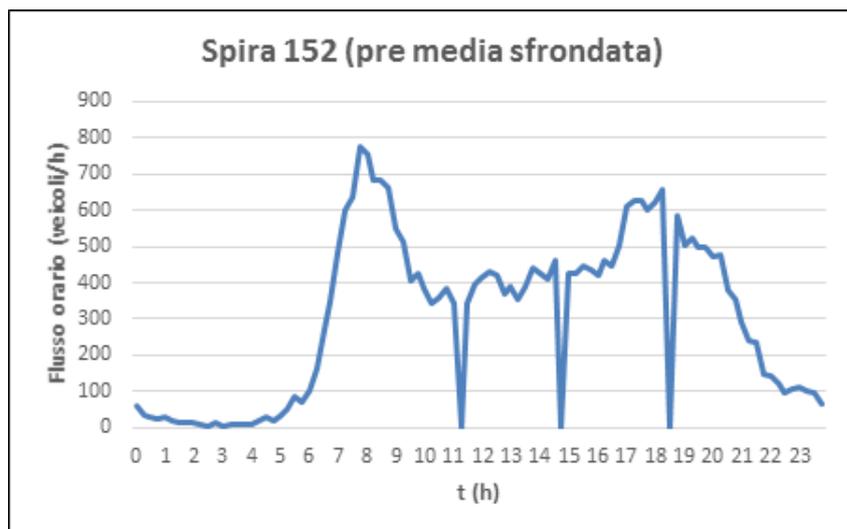
Per risolvere il problema dei dati mancanti e isolati è stata applicata una funzione di *media sfrondata* che dapprima raccoglie i valori dei quattro primi vicini del dato mancante, e dopo averli ordinati elimina il valore più basso (che è sempre lo 0 assegnato erroneamente al dato) ed il più alto, applicando

Figura 2.7: Frequenza dei valori nulli nelle spire.

al dato il valor medio dei tre rimanenti.

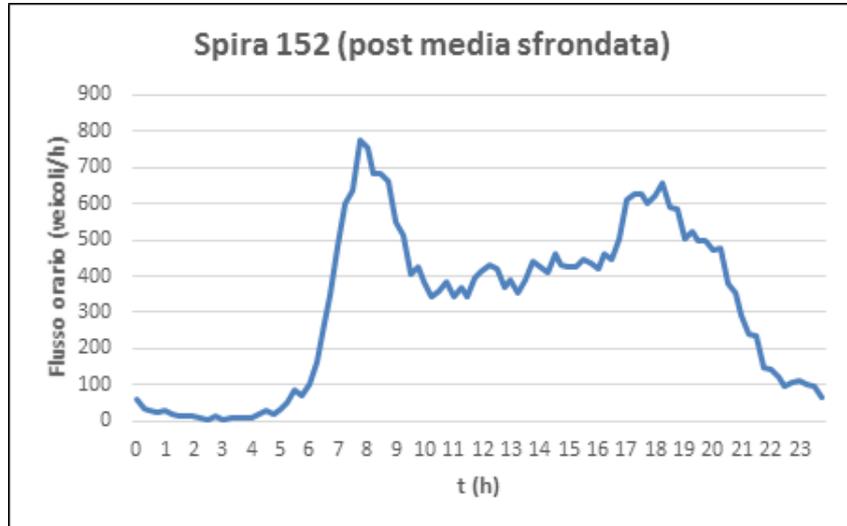
Quest'operazione si è applicata solo nell'intervallo diurno delle 07:00-22:00, poiché, come già affermato, per i dati in quello notturno l'assunzione che lo 0 fosse dovuto ad un errore della strumentazione era incerta.

Il risultato della media sfrondata applicata opportunamente permette di correggere una situazione come quella presentata in Fig. 2.8:

Figura 2.8: Una spira che presenta asimmetria fra le due direzioni di percorrenza.

Dopo l'applicazione del processo di ottiene infatti la Fig. 2.9:

Figura 2.9: Una spira che presenta asimmetria fra le due direzioni di percorrenza.



2.3.2 Dati GPS

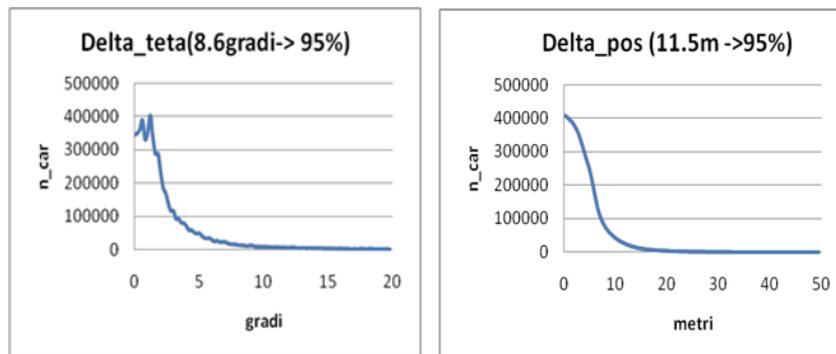
Il set di dati che in questo progetto verranno chiamati GPS[6] è un set raccolto da un dispositivo GPS/GSM chiamato *black box* che viene distribuito da diverse compagnie di assicurazioni a clienti interessati.

Il funzionamento tecnico di questi dispositivi si può schematizzare come segue:

- la prima misura, raccolta al momento dell'accensione, è raccolta in un buffer interno. Da quel momento il sistema effettua una misura ogni secondo;
- ogni misura che superi i due chilometri di distanza dall'ultima memorizzata viene inserita nel buffer;
- quando il buffer è riempito (dopo circa cinquanta misure), i dati che contiene sono inviati via GSM al server centrale di raccolta.

Chiaramente ogni dato è composto di una misura di posizionamento geografico. Nello studiare gli errori di posizionamento si è ottenuta una distribuzione di Cauchy con una varianza di circa 8 metri per la distanza e di circa 3 gradi per la direzione (nel caso di grafo rettilineo e con velocità v 10 Km/h). In Fig. 2.10 è mostrata la distribuzione degli errori nella distanza e nell'angolo per un campione di circa 5 milioni di dati.

Figura 2.10: Errori di posizionamento e di angolazione dei dispositivi GPS.



L'incertezza sui valori misurati è strettamente legata alla qualità della ricezione del segnale GPS al momento della misura: nelle figure è riportato il caso di massima qualità della ricezione GPS. L'errore sulla misura del tempo è in pratica trascurabile ai fini delle elaborazioni cui si sono sottoposti i dati (minore di 1 secondo). L'incertezza sulla misura della posizione è molto sensibile al problema della qualità del segnale. Quando la ricezione del segnale GPS è buona, come è stato osservato, l'errore di localizzazione è dell'ordine di 10 metri, valore che consente di posizionare adeguatamente i dati sulla rete stradale. L'incertezza sulle misure di velocità, direzione e distanza dalla misura precedente deriva da quella sulle misure di posizione da cui dipendono.

I dati sono sottoposti dapprima ad una pre-elaborazione atta a trasformare l'insieme grezzo delle misure in una sequenza ordinata e coerente di dati che possa essere usata per i passaggi successivi. Questa pre-elaborazione compie analisi di coerenza nella dinamica di ciascun veicolo registrato, ad esempio assicurandosi che le misure rispettino la sequenza corretta, per quanto riguarda lo stato di moto, di accensione \rightarrow movimento \rightarrow spegnimento. Successivamente

si procede ad uniformare le caratteristiche della sequenza di misure, eliminando le soste più brevi di un certo limite, o gli spostamenti spaziali più corti di un certo limite.

Quest'elaborazione iniziale ha inoltre permesso di associare un parametro di bontà ad ogni rilevazione, e per l'analisi si sono selezionate solo le misure con precisione maggiore.

I dati disponibili sono, dopo il processo di "pulitura":

Id veicolo	Codice identificativo del veicolo
Giorno	Data in cui è presa la misura (gg/mm/aa)
Ora	Ora in cui è presa la misura (hh:mm:ss)
Latitudine	Latitudine del veicolo in WGS84 (milionesimi di grado)
Longitudine	Longitudine del veicolo in WGS84 (milionesimi di grado)
Velocità	Velocità del veicolo in modulo (Km/h)
Direzione	Direzione di marcia (gradi, misurati dal Nord in senso orario) con una risoluzione di 2 gradi
Qualità del segnale	1 = assenza di segnale 2 = segnale scarso 3 = segnale buono
Stato del moto	0 = accensione motore 1 = marcia 2 = spegnimento motore 3 = ingresso nel periodo temporale considerato 4 = uscita dal periodo temporale considerato 5 = ingresso nella zona spaziale considerata 6 = uscita dalla zona spaziale considerata 7 = ingresso in una zona di dati affidabili 8 = uscita da una zona di dati affidabili
Distanza dal dato precedente	Distanza percorsa in moto dal dato precedente (metri)

E valgono le seguenti proprietà:

- non esistono misure con un intervallo temporale inferiore a 20 secondi;
- non esistono misure con un intervallo spaziale inferiore a 30 metri;
- non esistono intervalli di sosta di durata inferiore ai 30 secondi;
- non esistono misure di movimento di bassa qualità;
- la sequenza utilizzata tiene conto delle interazioni fra le varie operazioni e se riapplicata non ha effetti.

Posizionamento nella rete

Alcune misure di posizione, anche dopo la pre-elaborazione, sono affette da un errore che non permette di associarle immediatamente ad una strada della rete. L'assegnazione richiede in primo luogo la definizione di una distanza fra il punto misurato e le strade nel suo intorno.

Ad ogni strada con distanza minore di una certa soglia (50 metri) viene associata una probabilità di appartenenza tenendo anche conto dell'angolo tra la direzione del moto e quella dell'arco stradale. Se non si trova alcuna strada entro i 50 metri la misura viene scartata. Questo controllo ha portato a scartare il 3% delle misure. In prima approssimazione il dato può essere associato alla strada più vicina, oppure il processo può essere raffinato attribuendo al dato un'intera traiettoria, e quindi tenendo conto assieme di tutte le probabilità.

Calcolo delle traiettorie

Un ulteriore problema è stato quello di interpolare i punti nella rete stradale, cioè ricostruire la traiettoria realmente effettuata dal veicolo a partire dalle misure puntuali lungo il percorso.

Quest'operazione è complessa ed era già stata implementata in precedenza, ma per completezza verrà qui riassunta. Il percorso di ricostruzione delle traiettorie si articola nelle seguenti fasi:

- analisi di un intorno dei punti attribuiti dal sistema;
- attribuzione delle probabilità di appartenenza alle differenti strade;
- iterazione di questi passaggi fino ad arrivare al punto successivo;
- attribuzione di punteggi ai differenti percorsi rispetto a diversi fattori di merito, quali ad esempio percorrenza media della strada e storico dei percorsi dell'individuo.

Statistica dei dati GPS

Prima di ricostruire il flusso dei veicoli è necessario discutere l'affidabilità statistica della ricostruzione del numero totale di automobili a partire dal limitato numero di misure disponibili, poiché non si conosce con precisione la percentuale di auto con GPS rispetto alle auto totali.

Si consideri, ai fini di questa analisi, il seguente problema: si hanno $m + 1$ auto dotate di strumentazione GPS, e si vuole valutare quante auto senza GPS sono transitate tra loro. Se si considera, ad esempio, una penetrazione a pari a 1.34% si ottiene un valore medio di 75 auto ($= 1/a$). Se i passaggi di veicoli sono indipendenti fra loro ci si aspetta un processo Poissoniano[7] con una distribuzione nella forma:

$$P(n) = \alpha e^{-\alpha n} \quad (2.1)$$

Dove $P(n)$ descrive la probabilità di avere n auto comprese tra due dati noti.

E di conseguenza la distribuzione del numero di auto comprese tra $m + 1$ dati noti risulta essere:

$$P_m(n) = \frac{\alpha}{(m-1)!} (\alpha n)^{m-1} e^{-\alpha n} \quad (2.2)$$

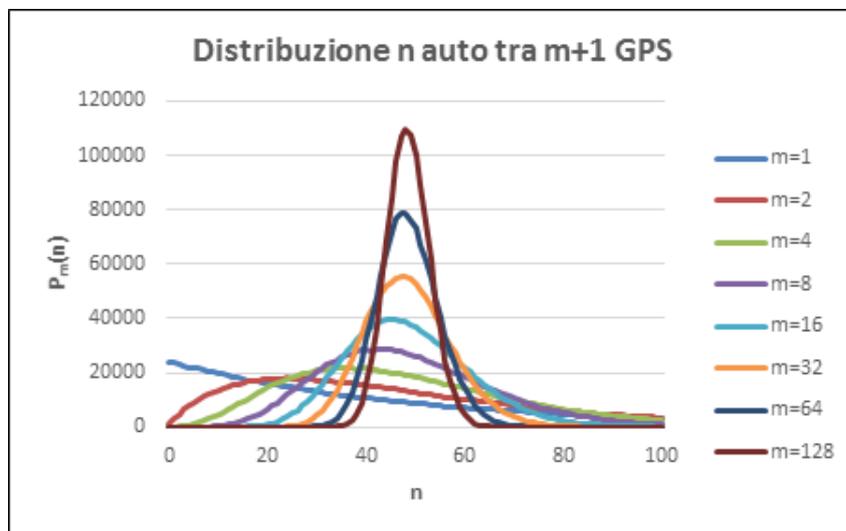
Nel derivare queste equazioni si è assunto che n sia una variabile continua e questo è corretto al limite per basse penetrazioni, mentre quando si hanno alte penetrazioni si dovrà considerare una distribuzione discreta.

Se ci si chiede quale sia il valor medio tra un'auto GPS e l'altra, date $m + 1$ tali auto, è necessario introdurre $j = n/m$ ottenendo che la distribuzione di probabilità è data da:

$$P_m(j) = \frac{\alpha m}{(m-1)!} (\alpha m j)^{m-1} e^{-\alpha m j} \quad (2.3)$$

Simulando il passaggio di un milione di veicoli, si ottiene per $P_m(n)$ la distribuzione in Fig. 2.11.

Figura 2.11: Distribuzione del numero di auto fra $m + 1$ auto GPS con penetrazione di 2%.



Come si può vedere queste distribuzioni sono asimmetriche con lo scarto quadratico medio sinistro maggiore di quello destro; questa differenza si riduce a circa un 10% per $m = 64$. La penetrazione considerata nel grafico mostrato è pari a 2%. Si ottengono i seguenti parametri statistici:

m	valor medio	sqm destro	sqm sinistro	Sqm
1	50	32.33	70.71	50.00
2	50	26.09	45.64	35.36
4	50	20.24	30.11	25.00
8	50	15.26	20.22	17.68
16	50	11.28	13.77	12.50
32	50	8.22	9.47	8.84
64	50	5.94	6.56	6.25

Se si considera lo scarto quadratico medio si vede che decade esattamente con la radice, ossia si può scrivere che il numero di auto fra $m + 1$ dati è il seguente:

$$n = \frac{m}{\alpha} \left(1 \pm \frac{1}{\sqrt{m}}\right) \quad (2.4)$$

Questo permette di valutare il flusso reale con il relativo errore, infatti, se si misura un flusso di m dati GPS in un intervallo di tempo T dalla (2.4) si ottiene:

$$\Phi = \frac{1}{\alpha} \frac{m}{T} \left(1 \pm \frac{1}{\sqrt{m}}\right) \quad (2.5)$$

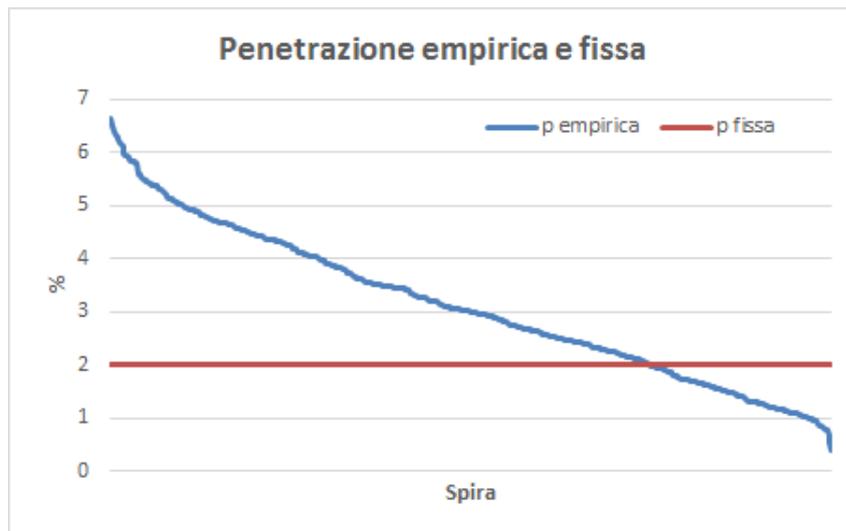
Il massimo di queste curve è sempre attorno a $n = 49$ il che implica che osservando un alto numero di eventi si ha comunque un valore medio di 49 auto non dotate di GPS tra due auto che ne sono provviste.

A partire da queste considerazioni si è scelto di escludere dalle analisi le strade in cui il flusso fosse basso, perché la ricostruzione a partire dalla penetrazione richiede che ci sia un certo numero di dati per essere valida. L'errore, se si decidesse di ricostruire anche le strade a basso flusso, aumenterebbe in grande misura andando ad inficiare in maniera molto evidente sulla ricostruzione dei flussi sull'intera regione.

2.4 La penetrazione dei dati GPS

Dopo aver analizzato separatamente i dati provenienti da spire e quelli da GPS è necessario studiare un parametro che ci permetta di rapportare questi dati di natura differente; questo valore è la già nominata *penetrazione* che definisce la percentuale di autoveicoli che possiedono una strumentazione GPS. Questo valore è puntuale, ed estendere una misura di penetrazione ad aree limitrofe ne diminuisce via via la precisione. Infatti, se si considera l'intera rete stradale italiana, si ottiene una penetrazione media del 2%, se invece si considera ogni postazione di spira, che raccoglie tutti i dati di transito di autoveicoli e si confronta questo valore con il valore di flusso ottenuto durante lo stesso periodo di tempo da dispositivi GPS, si osservano risultati molto differenti, come in Fig. 2.12.

Figura 2.12: Confronto visuale fra l'ipotesi di penetrazione uniforme e il calcolo empirico.



È quindi evidente che utilizzare per tutte le strade una penetrazione del 2% sarebbe stata una semplificazione troppo grossolana ed avrebbe rappresentato una grande fonte di errore nell'utilizzo dei dati GPS. Senza un valore sufficientemente accurato del parametro di penetrazione per ogni strada, non è possibile ricostruire i flussi nella regione, poiché non si sa quante auto effettive percorrano la strada, avendone solo una proporzione ignota.

Di fatto, il problema del calcolo di questo parametro è di difficile risoluzione per varie ragioni; i dispositivi GPS, ad esempio, non sono distribuiti omogeneamente fra le diverse fasce di età e di classe di reddito. Questa particolare problematica non è stata affrontata in questo lavoro, poiché le normative sulla privacy non consentono di diffondere a terzi le informazioni sull'età o sul reddito degli individui.

Vista la grande importanza del parametro di penetrazione ai fini di questo studio, si è cercato di determinarli con accuratezza, adottando diverse metodologie di calcolo. Si è inoltre tentato di classificare la regione in zone, nel tentativo di estendere i valori considerati buoni anche a strade in cui non è possibile confrontare direttamente i valori di flusso GPS con quelli di spira.

2.4.1 Calcolo in assenza di dati da spire

La prima metodologia di estensione a zone in cui non è possibile ottenere dati da spire (ad esempio altre città) è stata quella che si compone dei seguenti passaggi:

- valutazione del numero di auto totale in ogni comune della regione;
- attribuzione di un comune di appartenenza alle auto dotate di GPS;
- confronto tra i risultati per ottenere la penetrazione.

Per attribuire il numero delle auto totali ai differenti comuni di residenza sono state utilizzate le informazioni del database ACI delle auto immatricolate[8].

Per assegnare un comune di residenza ai dati GPS sono stati studiati, per ogni veicolo, gli intervalli tra misure vicine, distinguendo le fasi di sosta del veicolo da quelle di movimento; in questo modo si sono assegnati ai veicoli un *comune di residenza*, quello cioè dove compie le soste maggiori, ed un *comune di influenza*, cioè il comune in cui esercita la propria mobilità. Si è comunque osservato che vi sono distinzioni solo in una piccola percentuale dei veicoli.

Infine per calcolare i valori di penetrazione per i differenti comuni si sono utilizzati i dati dei GPS relativi alle zone di influenza, dividendo il numero di questa dati per il numero di auto immatricolate in quel comune.

È comunque importante notare che considerando come numero totale di auto le auto immatricolate si sottostima il valore di penetrazione, poiché non tutte le automobili immatricolate sono in circolazione; si è verificato che questo errore è comunque trascurabile.

La penetrazione calcolata in questo modo verrà chiamata *comunale*; questa modalità di calcolo del parametro di penetrazione è stata implementata per motivazioni “storiche”, traendola da progetti precedenti a quello qui presentato che non avevano a disposizione dati da spire, ma necessitavano della conoscenza di questa grandezza.

Estensione con matrice di punti

Successivamente si è scelto di raffinare il calcolo del coefficiente di penetrazione comunale, poiché attribuire un valore semplicemente legato al comune di appartenenza, ignorando la posizione geografica della strada all'interno di quel comune e le penetrazioni dei comuni adiacenti, era un'approssimazione probabilmente eccessiva.

È stata creata una griglia quadrata rappresentata da una matrice che è stata posizionata, in prima istanza, sul centro di ogni spira; si è poi applicata una media pesata delle diverse coordinate di questa matrice utilizzando come pesi il numero di punti nei vari comuni moltiplicati per i valori di penetrazione del comune stesso, così da ottenere un valore di penetrazione unico che tenesse conto dei comuni adiacenti in maniera proporzionale al volume di dati contenuti in ogni zona.

Per trovare la dimensione ideale della griglia sono state effettuate differenti simulazioni fino ad arrivare ad utilizzare il valore che minimizza la differenza tra penetrazione empirica ed il valore di penetrazione così calcolato, che verrà chiamato penetrazione *di matrice*.

Formalmente si è cercato empiricamente il raggio che minimizzasse

$$\sum_i |p_mat_i - p_es_i|. \quad (2.6)$$

In alternativa alla griglia quadrata scelta inizialmente è stato anche utilizzato un intorno circolare, scartando i punti della griglia distanti dal centro più del raggio scelto, per realizzare in pratica il concetto di un intorno composto di punti con distanza massima costante in ogni direzione.

Nella tabella si mostrano i risultati della ricerca dei minimi errori. Si noti che in entrambe le simulazioni è stata utilizzata una griglia 11x11.

Quadrato		Cerchio	
raggio (Km)	$\sum_i p_mat_i - p_es_i $	raggio (Km)	$\sum_i p_mat_i - p_es_i $
8	388.5	8	393.6
8.25	388.9	8.25	392.3
8.5	389.2	8.5	391.4
8.75	389.6	8.75	391.6
9	391.2	9	392.8
9.25	388.6	9.25	389.8
9.5	385.7	9.5	388.5
9.75	386.0	9.75	387.7
10	385.9	10	387.3
10.25	385.3	10.25	387.1
10.5	386.0	10.5	388.0
10.75	385.5	10.75	386.4
11	385.4	11	386.7
11.25	386.2	11.25	386.8
11.5	386.4	11.5	386.5
11.75	386.6	11.75	386.3
12	387.0	12	385.7
12.25	386.5	12.25	386.0
12.5	388.6	12.5	388.5
12.75	388.0	12.75	387.2
13	388.0	13	386.1
13.25	388.1	13.25	385.9
13.5	389.1	13.5	386.6
13.75	390.7	13.75	386.9
14	391.6	14	386.9

Con entrambe le metriche, comunque, si è osservato che il numero di punti scelto per comporre la griglia ha un'influenza limitata sul risultato, e utilizzare una matrice di 11x11 punti è già sufficiente; andare a 101x101 punti non migliora in maniera sostanziale la precisione (qualche punto percentuale) ed è

comunque computazionalmente pesante, rendendo proibitivo il calcolo su tutta la rete.

Bisogna soprattutto osservare che far variare il raggio della maschera da 8 a 14 Km cambia la differenza totale solo dell'1-2%. In sostanza si può concludere che un raggio di una decina di chilometri è una buona stima dell'influenza di un comune sulle penetrazioni della zona.

Questo metodo per calcolare valori di penetrazione è stato successivamente applicato ad ogni segmento di strada, centrando la griglia sul punto centrale di ogni segmento della poligonale che rappresenta la strada stessa. In questo modo è stato ottenuto un ulteriore metodo di propagazione d'informazione, affine ai metodi di convoluzione con maschera utilizzati, ad esempio, nella manipolazione di immagini digitali; in questo caso corrisponderebbe ad un metodo di smoothing con filtro di media[9].

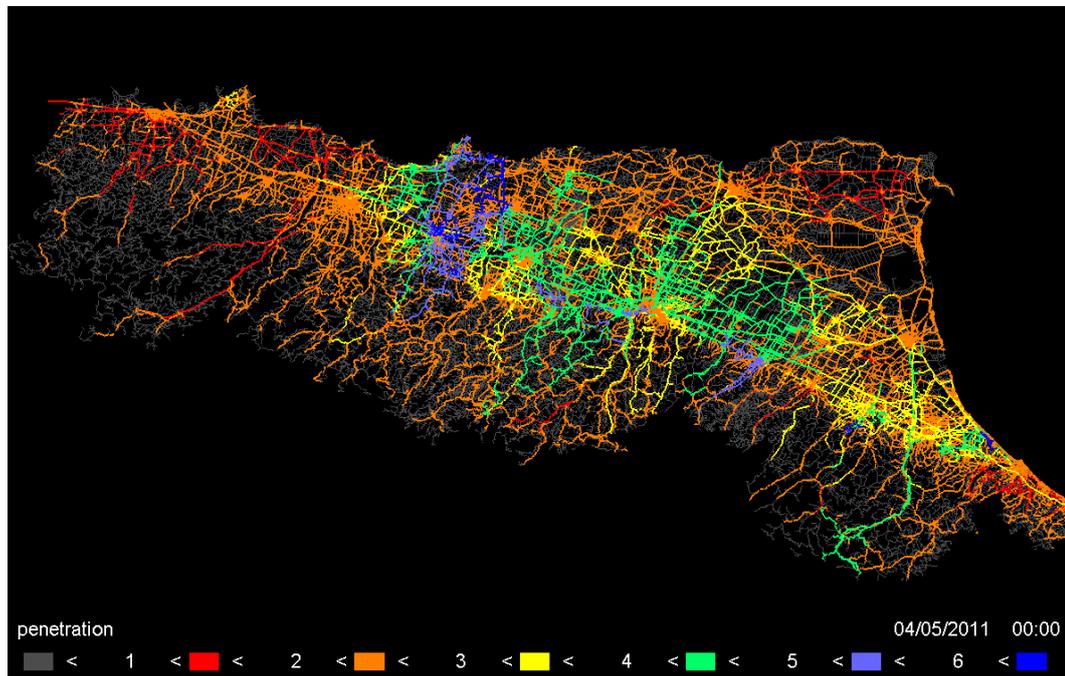
2.4.2 Calcolo con dati da spire

Successivamente queste stime di penetrazione sono state affiancate da quelle ottenute dal confronto dei dati GPS con quelli provenienti da spira. La penetrazione calcolata sulle spire semplicemente rapportando i due valori prende il nome di penetrazione *esatta*, perché in quelle postazioni si conosce esattamente il numero di tutte le auto transitate, comprese quelle dotate di GPS. Successivamente questi valori sono stati estesi su tutta l'area della regione tramite il metodo di propagazione dell'influenza delle spire alle strade, che verrà illustrato successivamente nel Capitolo 3. Questa varietà di penetrazione verrà chiamata penetrazione *empirica*.

Si è osservato che i valori di penetrazione empirica osservati presentano una buona coerenza spaziale, cioè che zone limitrofe hanno valori di penetrazione simili.

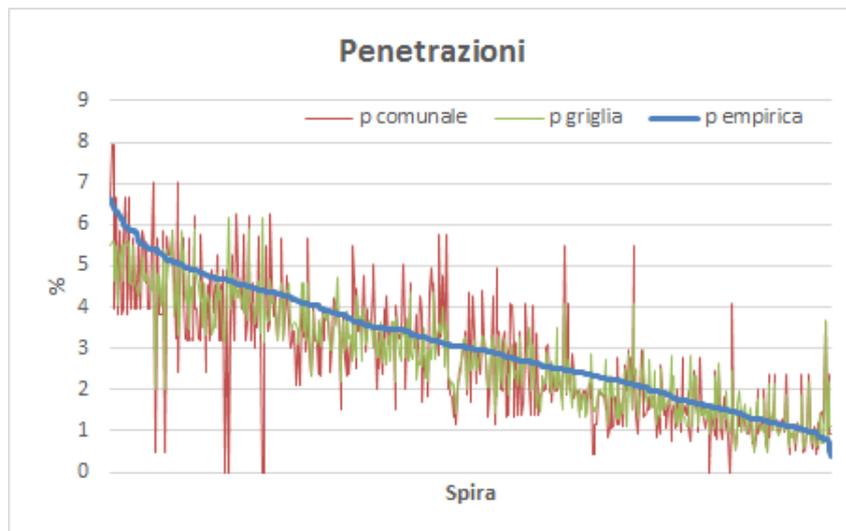
I valori di questa penetrazione sono mostrati in Fig. 2.13.

Figura 2.13: Rappresentazione geografica della penetrazione empirica dei dati GPS.



2.4.3 Conclusioni

Figura 2.14: I tre tipi di penetrazione a confronto.



Nel grafico in Fig. 2.14 è mostrato un confronto fra i tre tipi di penetrazione

calcolati, ordinando le spire in base al valore di penetrazione esatta.

Nonostante che la penetrazione calcolata con la matrice sia una valida alternativa a quella empirica, si è comunque scelto di utilizzare quest'ultima nelle simulazioni. La motivazione è duplice:

- utilizzando la penetrazione a matrice si ottengono risultati peggiori che utilizzando la penetrazione empirica, che sfrutta i dati GPS;
- la definizione di penetrazione empirica è più coerente con le metodologie di calcolo utilizzate nella simulazione, in cui i valori spaziali delle grandezze vengono estesi a partire da pochi dati locali.

Capitolo 3

Algoritmi di propagazione

3.1 Introduzione

Prima di iniziare a descrivere il modello sviluppato, è importante ricordare i dati a disposizione, in quanto il modello è stato realizzato per ricostruire i flussi su tutta la rete della regione, utilizzando al meglio le informazioni disponibili.

I dati disponibili sono quelli provenienti dalle spire e quelli dai veicoli con GPS:

- quelli che provengono dalle spire sono dati sparsi, limitati in numero ma accurati;
- quelli provenienti dai GPS sono dati numerosi e diffusi, ma spazialmente disomogenei.

Inoltre si ha disposizione una cartografia digitale della rete regionale, quindi una conoscenza della topologia del network, e si hanno informazioni sulle caratteristiche delle varie strade, quali la tipologia stradale (strada regionale, provinciale, etc...), l'estensione in lunghezza, il numero di corsie, la velocità di crociera media e quella media in assenza di traffico (detta *free speed*).

I modelli sviluppati non hanno la pretesa di simulare perfettamente e interamente la rete della regione Emilia Romagna: per ragioni già esposte si

è fatta una selezione di strade con valore di percorrenza relativamente alto, all'incirca di 1000 veicoli al giorno. Inoltre sono state escluse strade sterrate o strade private.

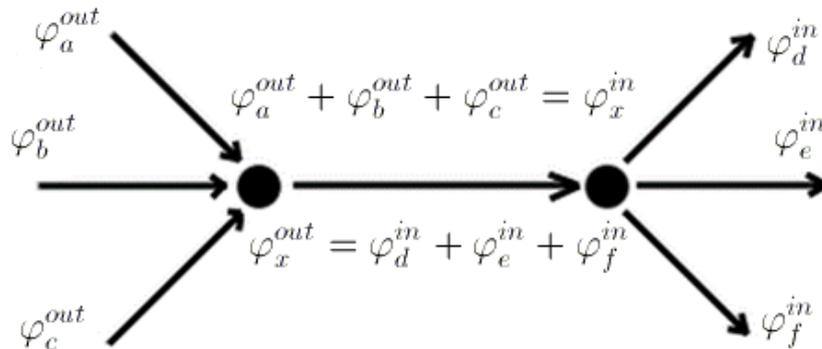
3.2 Trattazione teorica

Prima di esporre le metodologie pratiche adottate nei modelli, è opportuno presentare una trattazione teorica dei concetti alla base dei modelli.

Alla base del sistema vi è un principio di continuità attraverso i nodi della grandezza flusso[10]. Per un network, la conservazione si traduce semplicemente nella considerazione che, in ogni nodo, ad ogni istante, il flusso uscente coincide con la somma dei flussi entranti.

La Fig. 3.1 mostra un esempio di conservazione del flusso: tre strade confluiscono in una, il cui flusso equivale alla loro somma.

Figura 3.1: Rappresentazione schematica della conservazione del flusso.



Indicando con $\varphi_i^{(in)}$ e $\varphi_i^{(out)}$ rispettivamente il flusso in entrata sull'arco i -esimo e in uscita da esso, si ha per l'arco j

$$\varphi_a^{(out)} + \varphi_b^{(out)} + \varphi_c^{(out)} = \varphi_x^{(in)} \quad (3.1)$$

Spostando la propria attenzione sulle strade uscenti dall'arco j -esimo, ed introducendo la probabilità di svolta dall'arco j all'arco k (il c_{jk}), si ottiene

$$\begin{cases} \varphi_d^{(in)} = c_{xd} \varphi_x^{(out)} \\ \varphi_e^{(in)} = c_{xe} \varphi_x^{(out)} \\ \varphi_f^{(in)} = c_{xf} \varphi_x^{(out)} \end{cases} \quad (3.2)$$

Dove le probabilità di svolta sommano a 1: $c_{xd} + c_{xe} + c_{xf} = 1$.

Durante l'arco della giornata, come anticipato nel Capitolo 2, i valori di flusso raggiungono i massimi nelle ore di picco diurne e quasi si annullano durante le ore notturne, mentre le probabilità di svolta sono molto più stabili nel tempo.

3.3 Modelli

I modelli realizzati si basano sulla conoscenza esatta dei flussi sulle strade provviste di spira.

Il primo modello sviluppato propaga l'informazione che origina nelle spire alle strade vicine, utilizzando la ricostruzione del campo di velocità e delle probabilità di svolta ottenute dai dati GPS.

In pratica si parte dal valore di flusso su una spira e seguendo la rete si definiscono i flussi sulle strade a valle e si continua fino a che i flussi non diventano inferiori a una certa soglia minima, oppure entrano in una strada con un'altra spira. Questa operazione si fa per tutte le spire ed il flusso risulta la somma di tutti i flussi originati dalle singole spire (si è verificato che la presenza di loop non altera significativamente il risultato).

Per definire i flussi a valle si utilizzano le probabilità di svolta calcolate dai GPS, introducendo anche un ritardo temporale fra il valore del flusso calcolato e quello misurato dalla spira di origine calcolato dalla velocità media delle singole strade, anche questa valutata da dati GPS.

Questo metodo diffonde bene i flussi, mentre non descrive in alcun modo

le sorgenti tra una spira e l'altra ed è stato riscontrato che sottostima pesantemente i flussi presenti nella rete. In pratica è come se si considerassero le spire come uniche sorgenti, e quindi si risolvono bene i flussi nel loro intorno, mentre lontano da esse non arriva alcuna informazione rilevante. Inoltre la scelta di escludere le strade meno importanti dalla rete fa sì che, quando un veicolo entra in una di queste strade "sparisca" dal sistema portando ad una diminuzione del numero di veicoli e quindi, in mancanza di analoghe sorgenti che bilancino questi *pozzi*, il sistema perde veicoli in continuazione, sottostimando i flussi.

Per superare questi problemi, si è sviluppato un secondo modello seguendo un percorso alternativo, quello cioè di prendere in esame i rapporti fra i flussi sulle strade dotate di spire e i flussi sulle strade vicine.

Questo secondo metodo si basa sulle seguenti osservazioni:

- nelle ricostruzioni dei flussi dai dati GPS si è riscontrato che i rapporti fra i flussi di strade vicine tendono a mantenersi costanti all'interno delle due situazioni di picco di traffico a cui siamo interessati;
- si è appurato, nei flussi misurati dalle spire e nelle ricostruzioni dei flussi a partire dai dati GPS, che il sistema tende a essere isocrono.

Le spire quindi non sono più considerate delle sorgenti di segnale, ma come punti di riferimento per fissare istante per istante le ampiezze dei flussi delle strade circostanti, utilizzando i valori dei rapporti calcolati a partire dai dati GPS per le due situazioni di picco di traffico.

Se si avesse un'unica spira, in questo modo, essa governerebbe il flusso su tutta la regione con il suo valore. In presenza di più spire si è introdotta una media pesata che tiene conto della distanza dalla spira.

In questa prospettiva i flussi in una determinata strada saranno espressi come combinazione lineare dei flussi nelle strade vicine, le quali saranno ricorsivamente espresse nello stesso modo, risalendo fino alle spire come fonte di

informazione.

$$\forall j \quad \varphi_j = \sum_i w_{ij} \varphi_i \quad (3.3)$$

Il tutto si traduce nell'articolazione dell'algoritmo in tre passaggi:

1. nel primo si calcolano tutti i flussi sulle strade della rete nel picco di traffico considerato, ed essi sono utilizzati per calcolare i rapporti tra le strade senza spira e quelle con spira;
2. nel secondo si diramano dei pacchetti a partire dalle spire per attribuire alle varie strade senza spira le spire da cui dipendono ed il relativo peso (i w_{ij} dell'equazione sopra);
3. nel terzo ed ultimo passaggio si attribuisce il flusso istante per istante utilizzando i dati delle spire, i rapporti ed i pesi calcolati precedentemente.

I primi due passi vengono calcolati una volta sola all'inizio della simulazione, mentre il terzo viene calcolato in real time.

Dal punto di vista della computazione questo rende semplice l'utilizzo del programma per tempi arbitrari, poiché basta continuare a fornire in input i dati da spira per ottenere una simulazione continua dei flussi.

3.3.1 Calcolo dei pesi

Si sono sperimentati fattori peso di diversa natura, constatando che quelli che portano a risultati migliori sono quelli basati sulla probabilità di svolta e quelli calcolati a partire dal tempo di percorrenza delle strade.

Questo secondo metodo ha la proprietà di tenere conto dei pozzi e delle sorgenti del sistema in maniera implicita, poiché il sistema riscalda autonomamente in base ai flussi della zona; nel modello si utilizzano solo i flussi in uscita dalle strade, diversi dai flussi in entrata (che possono essere calcolati, si ricorda, come somma dei flussi uscenti dalle strade che si immettono) e la dif-

ferenza permette di calcolare la quantità di veicoli che entrano o escono lungo la strada.

In particolare in questo modo le auto che finiscono in strade escluse dal modello non spariscono permanentemente, ma possono rientrare nel sistema quando ritornano su una strada con un flusso abbastanza grande; ossia si considerano queste strade con flussi bassi come pozzi o sorgenti, ed il loro effetto viene implicitamente preso in considerazione dal modello.

Per ottenere i valori dei pesi, vale a dire dell'influenza delle strade vicine sulla spira principale, si sono utilizzati due differenti metodi:

- il primo basato sul tempo di propagazione dell'informazione, calcolato utilizzando i tempi medi di percorrenza delle strade;

$$w_{ji} = \frac{1}{t_{trip_{ij}}} \quad (3.4)$$

- il secondo invece basato sulle considerazioni alla base del primo modello, cioè la quantità di flusso derivata dalla singola spira, calcolato utilizzando le probabilità di svolta nelle varie strade.

$$w_{ji} = p_{svolta_{ij}} = \frac{n_j^{(out)}}{n_i^{(out)}} \quad (3.5)$$

È importante sottolineare che quando si parla di ritardo nel passaggio del segnale, si assume che esso sia dovuto solamente al tempo di percorrenza delle strade, e le svolte sono considerate istantanee. I possibili ritardi dovuti al passaggio da una strada alla successiva, dovuti a precedenza, semafori o altro, sono inglobati nel tempo di percorrenza della strada stessa.

I valori dei pesi vengono normalizzati in modo che $\sum_j w_{ji} = 1$.

Viene successivamente calcolato il valore di penetrazione empirica per ogni strada come:

$$p_{gps_i} = \sum_j w_{ji} p_{gps_j} \quad (3.6)$$

Poiché i pesi sono già normalizzati, si ottiene un valore di penetrazione per la strada equivalente alla media delle penetrazioni delle sue strade influenti.

3.3.2 Propagazione dei flussi

Una volta che i pesi sono stati calcolati come mostrato sopra, la simulazione deve propagare il segnale che, in tempo reale, viene fornito al programma sotto forma di flussi orari istantanei nelle strade con spira.

Anche per la propagazione si sono sviluppati diversi algoritmi, basati su un insieme di parametri via via più completi.

- Il primo metodo non associa dei ritardi alle propagazioni, ignorando il tempo di transito e assegnando semplicemente ad ogni strada la somma dei segnali delle strade con spira che la influenzano. In formule:

$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t) \quad (3.7)$$

- Il secondo schema è uguale al primo, eccetto per l'inserimento dei tempi di transito nel parametro temporale:

$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t - t_{trip_{ij}}) \quad (3.8)$$

Si è verificato che inserire questi ritardi nella propagazione non altera sostanzialmente i risultati, quindi si è scelto di non utilizzarli nelle propagazioni, ed il secondo schema quindi non viene successivamente utilizzato nelle varie simulazioni.

L'ultimo metodo sviluppato, infatti, non tiene conto dei ritardi. In questo schema vengono presi in considerazione i rapporti fra i flussi delle strade vicine e anche i valori di penetrazione. Questo è l'algoritmo più completo e quello che produce errori minori all'atto della validazione del modello.

Formalmente ad ogni strada viene attribuito un flusso calcolato con

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p-gps_i}{n_j^{(out)}/p-gps_j} \varphi_j(t) \quad (3.9)$$

3.3.3 Validazione del modello

Ogni modello deve essere accompagnato da una fase di verifica della propria validità.

Nel presente lavoro, la validazione è basata sull'attribuzione di un valore di errore ai risultati ottenuti; poiché lo scopo di questo lavoro è la ricostruzione del campo di flusso regionale, l'errore viene calcolato in base a quanto queste ricostruzioni si discostino dalla realtà.

Più specificamente, l'errore è calcolato da un confronto fra i dati reali, provenienti dalle spire, e la ricostruzione sulle stesse spire fatta dal modello; questa ricostruzione è calcolata come se la strada non avesse la spira, cioè considerando che il flusso su questa strada sia il risultato dell'influenza delle altre spire.

I dettagli di questo processo di validazione non saranno qui discussi, poiché il responsabile di questo argomento non è l'autore di questa tesi. Nondimeno, i risultati di queste validazioni verranno presentati nel Capitolo 5, dove si mostrerà quali algoritmi diano buoni risultati e quali siano meno validi.

3.4 Rappresentazione schematica del modello

In questa sezione viene presentata in forma visuale una descrizione schematica della struttura computazionale del modello.

Figura 3.2: Il flowchart principale.



Figura 3.3: La sezione di Input.

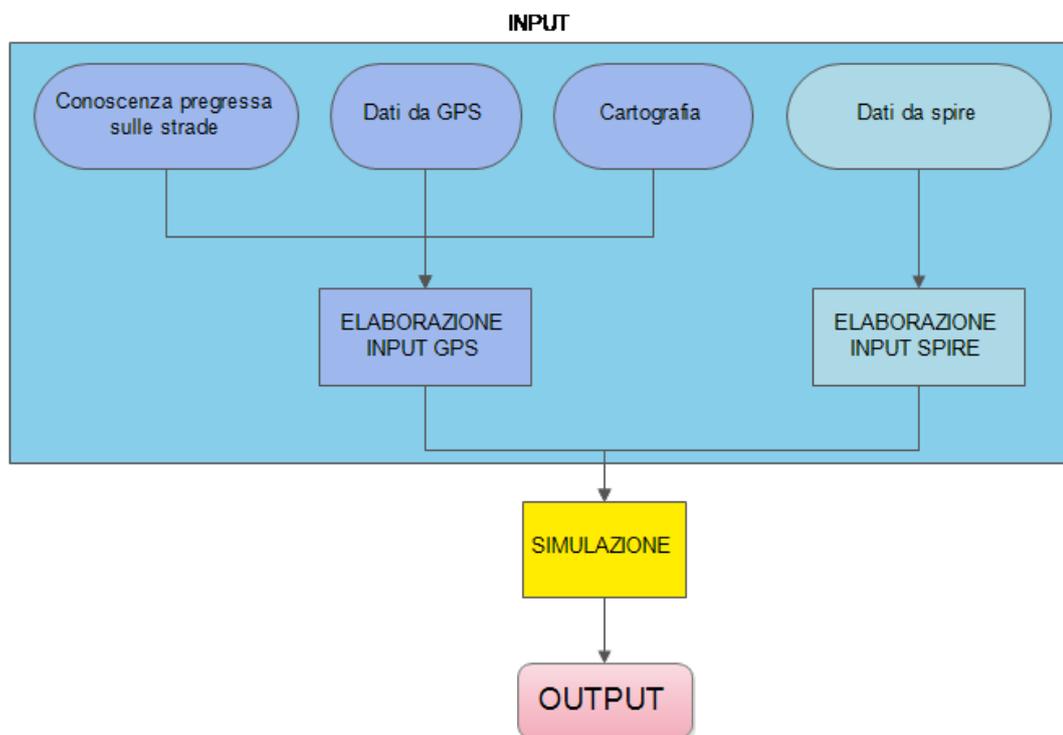


Figura 3.4: La simulazione e calcolo delle ricostruzioni.

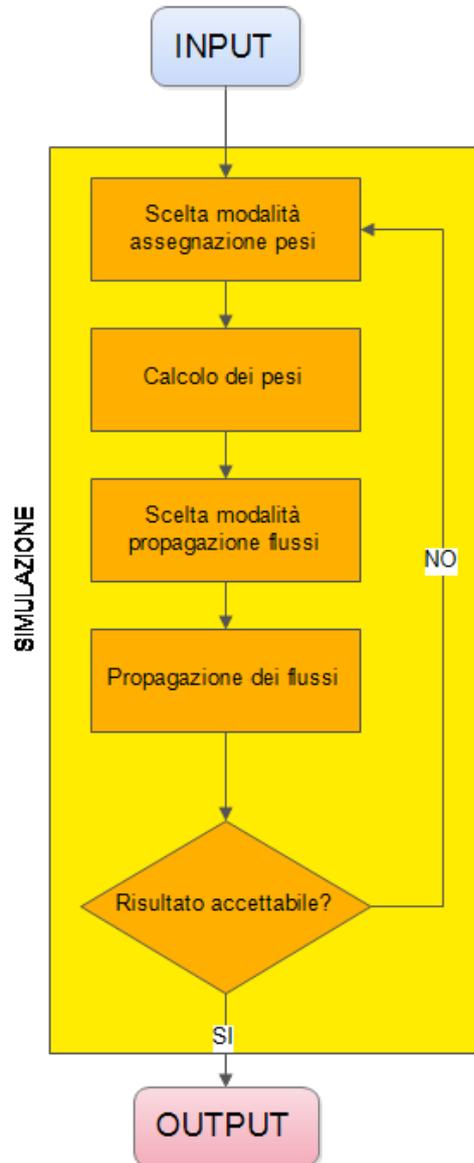
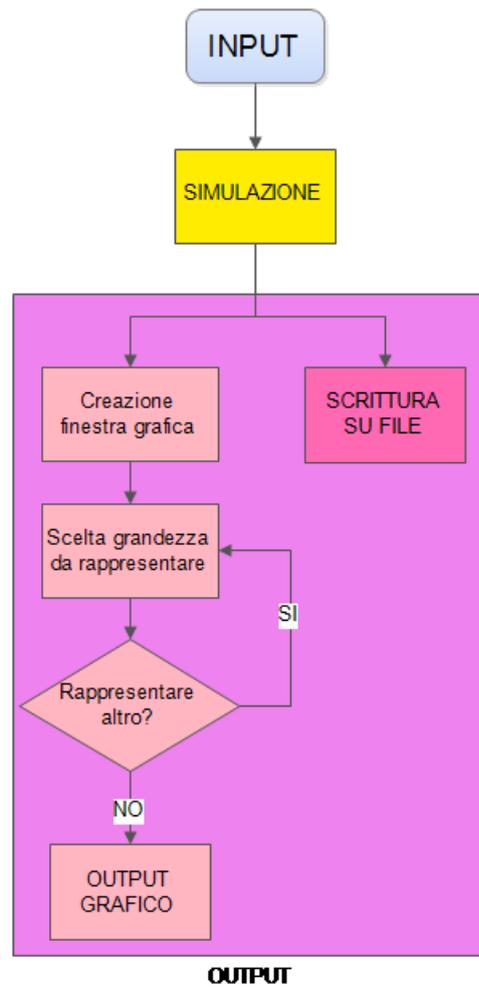


Figura 3.5: La sezione di output.



Capitolo 4

Analisi dati

In questo capitolo è illustrata la parte del lavoro di tesi in cui si è tentato di effettuare un'analisi diretta dei dati, prescindendo momentaneamente dai risultati del modello. In realtà queste analisi, mirate alla ricerca di comportamenti e relazioni interessanti fra dati, e dell'interpretazione delle variazioni delle grandezze, portano ad una conoscenza che è un fattore integrante per migliorare ed ottimizzare il modello.

Le analisi qui descritte interessano i flussi giornalieri mediati sulle giornate feriali del mese di maggio 2011, frutto della selezione descritta nel Capitolo 2.

4.1 Correlazioni fra spire

Molte delle analisi riguardano informazioni sulle relazioni fra spire vicine, ricavate da posizione geografica e influenza reciproca nella struttura di calcolo del modello; questa influenza deriva dalla vicinanza e dai rapporti fra flussi delle strade che congiungono gli apparati.

Un primo studio è stato condotto per trovare relazioni fra tutte le possibili coppie di spire e per evidenziare la presenza di una componente isocrona e diffusa degli andamenti giornalieri. Come indice di similarità fra spire si è utilizzata la correlazione statistica.

Si rimanda alla tesi del collega per queste analisi.

4.2 Distinzione in fascia mattutina e pomeridiana

In una seconda analisi del problema ci si è chiesto come la situazione cambiasse tra la fascia mattutina e pomeridiana nel panorama della mobilità dell'Emilia Romagna. Questa domanda è nata dall'ipotesi che nelle diverse ore della giornata ci fossero comportamenti differenti e quindi l'idea era quella di migliorare l'intero modello di calcolo su queste basi. Ad esempio, diverse zone della città possono diventare centri di attrazione ad alcuni orari, o alcune strade possono venire percorse in maniera differente in diverse ore della giornata. Peraltro, è già stato mostrato che si sono osservate delle differenze di transito nei profili giornalieri anche nelle due corsie di alcune spire, e anche questa caratteristica ha suggerito che un'analisi di questo tipo potesse rivelarsi utile.

In particolare si sono studiati, dal punto di vista della divisione in fasce, tre parametri:

- le penetrazioni;
- le probabilità di svolta tra le differenti strade;
- il parametro di validità del modello, vale a dire gli errori commessi nella simulazione dei flussi.

I risultati di queste divisioni sono trattate in maniera esauriente nella tesi dell'altro laureando.

4.3 Parametri caratteristici delle spire

Come sintesi dei vari studi effettuati sulle spire, si è tentato di definire dei parametri, associati ad ogni spira, che ne raccogliessero e sintetizzassero

diverse caratteristiche nel contesto del modello sviluppato.

Si sono selezionate tre tipologie di informazione ritenute tipiche della natura degli apparati, arrivando alla definizione di tre indici distinti:

- la *bontà*, che è una misura di come e quanto la spira viene utilizzata all'interno del sistema;
- l'*influenza spaziale*, che indica l'estensione geografica della zona di influenza della spira;
- l'*influenza di flusso*, che rappresenta un indice dell'entità dei flussi "trasmessi" dalla spira.

Questi indici permettono appunto di riassumere le caratteristiche delle spire all'interno dell'ambiente in cui sono utilizzate.

Il primo, sebbene sia una proprietà che descrive una spira, è un indice strettamente dipendente da altre spire; i due indici d'influenza sono invece grandezze indipendenti dalla presenza di altre apparati nel sistema, e nel calcolarle si tiene in conto solo della spira e della struttura del suo intorno.

Per le definizioni e una trattazione completa di questi parametri, si veda la tesi del collega.

Capitolo 5

Risultati

In questo capitolo vengono mostrati i risultati delle simulazioni. Come già esposto nel capitolo sugli algoritmi di propagazione, si sono sviluppati diversi schemi per estendere le informazioni di flusso alle zone non dotate di spire. Si riassumono qui i diversi metodi di propagazione adottati:

- pesi calcolati in base ai tempi di viaggio, propagazione tenendo conto dei rapporti di flusso;
- pesi calcolati in base ai tempi di viaggio, propagazione senza tener conto dei rapporti di flusso;
- pesi calcolati dalle probabilità di svolta, propagazione senza rapporti di flusso;
- pesi calcolati dalle probabilità di svolta, propagazione con rapporti di flusso.

Per stabilire la validità del modello sviluppato si osserva, per ogni spira, la ricostruzione che il modello fa del flusso sulla spira a partire dai suoi vicini e ignorando la misura reale sulla spira stessa e la direzione opposta.

La formula utilizzata per il confronto effettivo è la somma

$$E = \sum_i \int_{t_0}^T \epsilon_i(t) dt \quad (5.1)$$

dove $\epsilon_i(t) = |\varphi_i^{(misur)}(t) - \varphi_i^{(calcol)}(t)|$ è l'errore istantaneo della spira i .

La somma risultante è poi divisa per il numero di misure del periodo considerato, che nel caso di una giornata sono 96, e per il numero N di spire, ottenendo un errore medio.

$$\bar{E} = \frac{E}{96 N} \quad (5.2)$$

Questo capitolo illustra che utilizzare diversi criteri cambia i risultati del computo di questo parametro di errore del modello.

Si mostreranno di seguito, per i diversi metodi, i risultati dell'estensione dei flussi a tutte le strade che sono mantenute nel modello, escludendo quelle eliminate per carenza di informazioni o per la loro natura poco interessante.

È già stato detto che la divisione in fasce orarie della giornata produce un miglioramento dell'errore finale di circa 7-8%. Per questa ragione gli errori mostrati in questo capitolo, ove non altrimenti specificato, saranno quelli che risultano dall'attuazione di questa suddivisione.

È opportuno ricordare anche che le autostrade, sebbene siano ovviamente strade molto trafficate, sono state escluse dal modello non essendo monitorate dalle spire della Regione, quindi nelle mappe mostrate non appariranno.

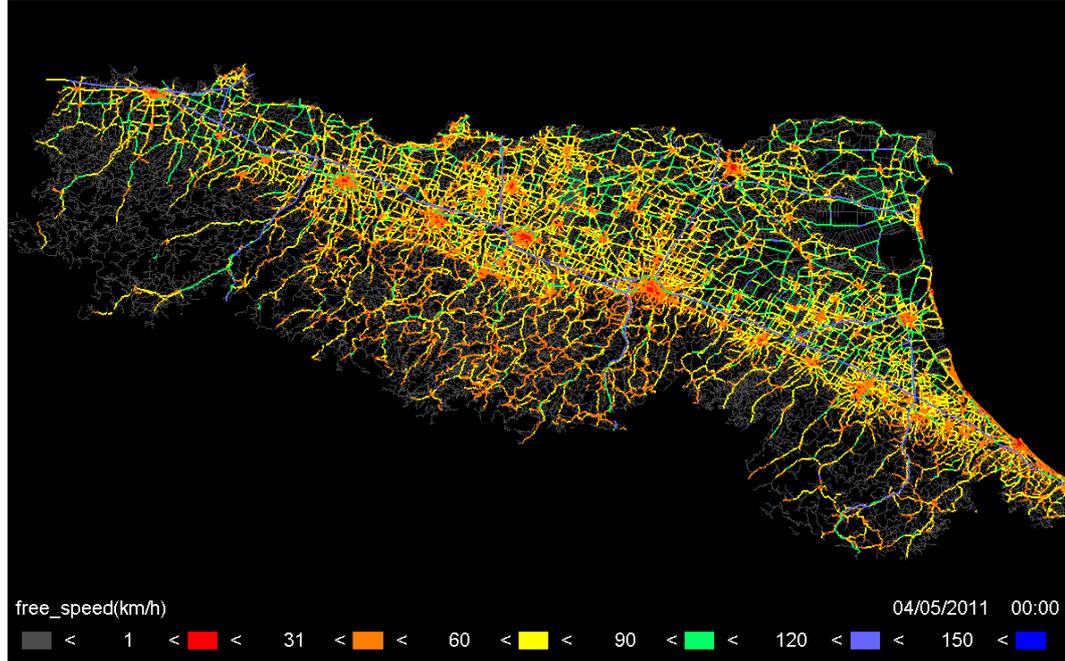
Col programma sviluppato è inoltre possibile osservare i valori di velocità libera media, la *free speed*, che come già detto sono ricavati direttamente dai dati GPS, ed è definita come l'85° percentile del range di velocità misurato in ogni strada. Si mostrerà quindi anche un'immagine della *free speed*.

Per questa grandezza le figure non variano nel tempo poiché i valori sono medie calcolate a partire da tutte le velocità misurate nel mese, secondo il metodo esposto in precedenza.

5.1 Pesi con tempi di viaggio, propagazione con rapporti di flusso

In Fig. 5.1 è mostrata questa grandezza sulla mappa della regione. Si osserva che nelle città la velocità rimane bassa, mentre nelle strade ad alta percorrenza è ovviamente elevata.

Figura 5.1: Rappresentazione su mappa delle *free speed*.



5.1 Pesi con tempi di viaggio, propagazione con rapporti di flusso

Utilizzando questi parametri, i pesi sono calcolati come

$$w_{ji} = \frac{1}{t_{trip_{ij}}} \quad (5.3)$$

E la propagazione avviene con

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p_{gps_i}}{n_j^{(out)}/p_{gps_j}} \varphi_j(t) \quad (5.4)$$

5.1 Pesì con tempi di viaggio, propagazione con rapporti di flusso 55

Il grafico rappresentato in Fig. 5.2 mostra l'errore medio delle spire, ordinandole in base all'errore stesso in maniera decrescente; in Fig. 5.3 è un istogramma che mostra la distribuzione dell'errore.

Il periodo su cui questa simulazione è stata compiuta è l'intera giornata del 4 maggio 2011.

Figura 5.2: Ranking dell'Errore con giornata divisa, pesi con tempi, propagazione con rapporti.



L'istogramma ed il grafico mostrano che l'errore è mediamente basso. Solo pochissime spire hanno errore al di sopra dei 150 veicoli/h, e l'errore medio E_m vale 54.2 veicoli/h.

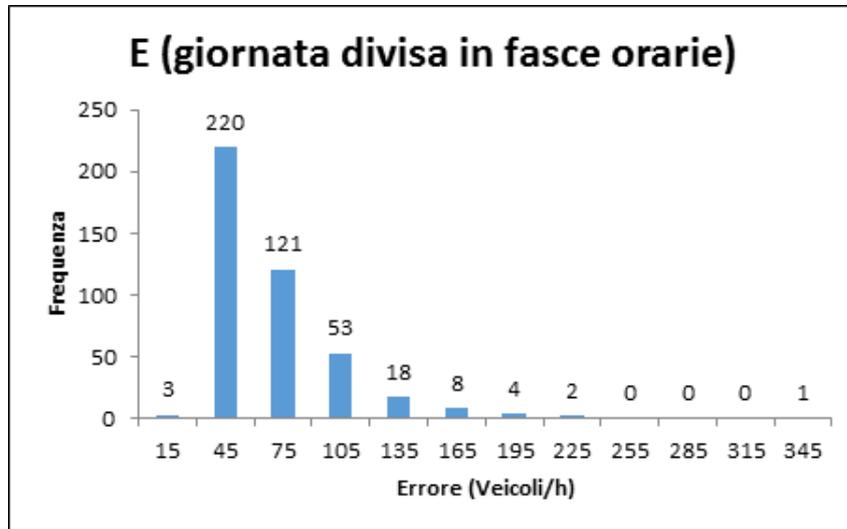
Le poche spire problematiche, si ricorda, sono attribuite al fatto che il loro grande flusso non permette dal punto di vista matematico un'ottima ricostruzione.

In Fig. 5.4 viene mostrato l'errore rappresentato cromaticamente sulla mappa regionale, per dare un'idea delle zone ricostruite meglio e peggio da questo metodo.

È opportuno sottolineare nuovamente il fatto che questo parametro di errore è l'errore massimo commesso su una spira se essa non ci fosse, e nell'estensione dei flussi alle zone attorno alle spire l'errore, per le ragioni esposte

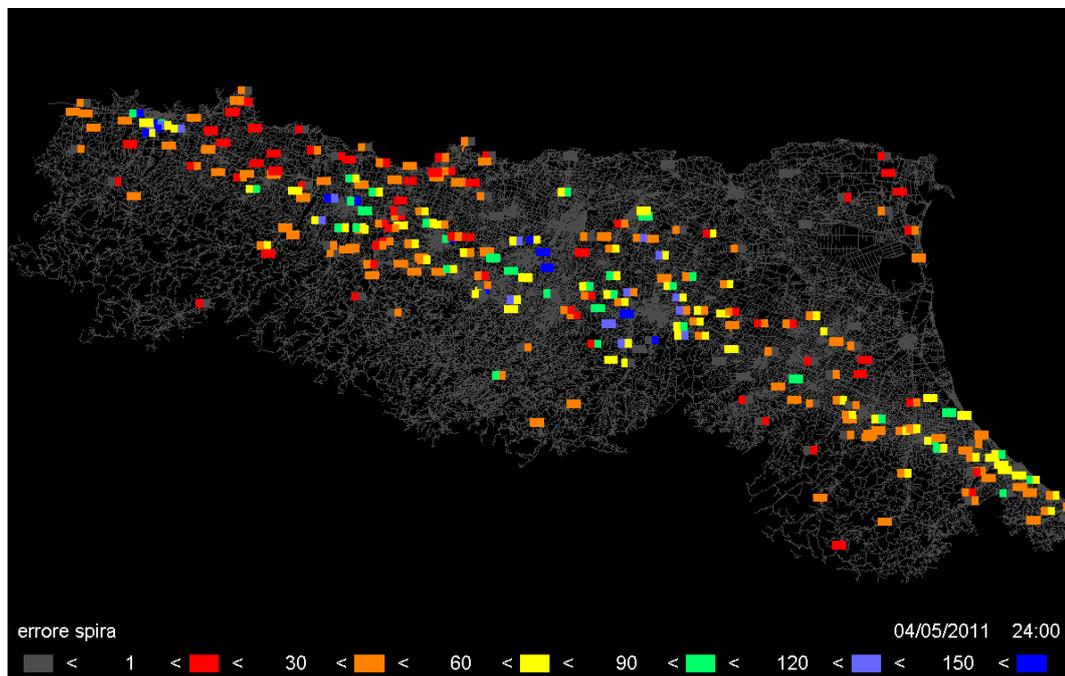
5.1 Pesì con tempi di viaggio, propagazione con rapporti di flusso 56

Figura 5.3: Distribuzione dell'Errore con giornata divisa, pesi con tempi, propagazione con rapporti.



nel Capitolo 3, sarà minore del valore espresso.

Figura 5.4: Gli errori di ricostruzione dei flussi in rappresentazione geografica, con lo schema di calcolo più valido.



5.1 Pesi con tempi di viaggio, propagazione con rapporti di flusso

Si noti la convenzione di rappresentare ogni spira con due rettangoli associati alle due direzioni di marcia. Ovviamente, non è scontato che i due versi assumano lo stesso errore, pur essendo esattamente nello stesso punto geografico.

È possibile osservare anche dalla mappa che le zone in cui l'errore risulta alto sono poche, e queste zone sono nelle immediate vicinanze di città relativamente importanti come Bologna, Modena, Parma e Piacenza; la ricostruzione migliora spostandosi anche di poco dai centri urbani.

Queste considerazioni suggeriscono che gli errori assoluti siano grandi dove i flussi sono maggiori, come nei centri urbani. Questa questione è discussa in dettaglio nella tesi dell'altro laureando.

5.2 Pesi con tempi di viaggio, propagazione senza rapporti di flusso

In questo caso i pesi sono calcolati come:

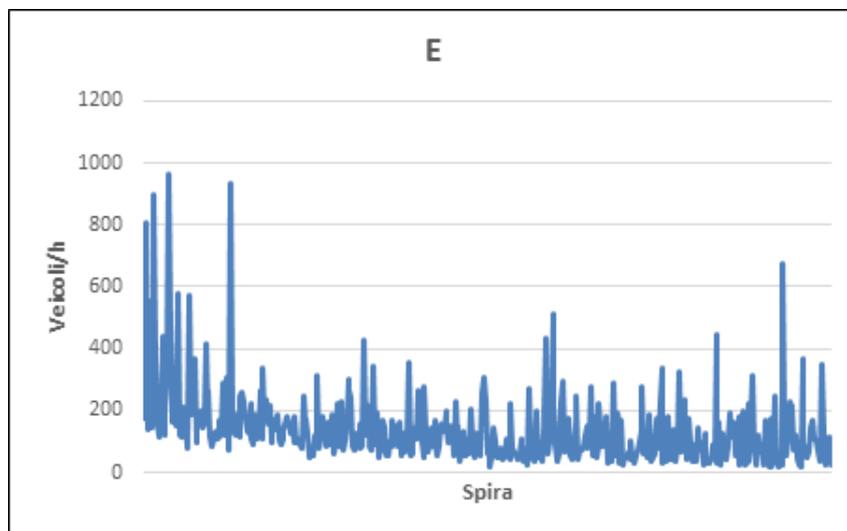
$$w_{ji} = \frac{1}{t_{trip_{ij}}} \quad (5.5)$$

e la propagazione avviene con

$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t) \quad (5.6)$$

Analogamente al paragrafo precedente, si propongono (Figg. 5.5 e 5.6) i due grafici di questo errore. In questo caso, tuttavia, la divisione in fasce della giornata non può essere applicata al calcolo dell'errore, poiché nessuna variabile che contribuisce al calcolo è dipendente da valori medi giornalieri.

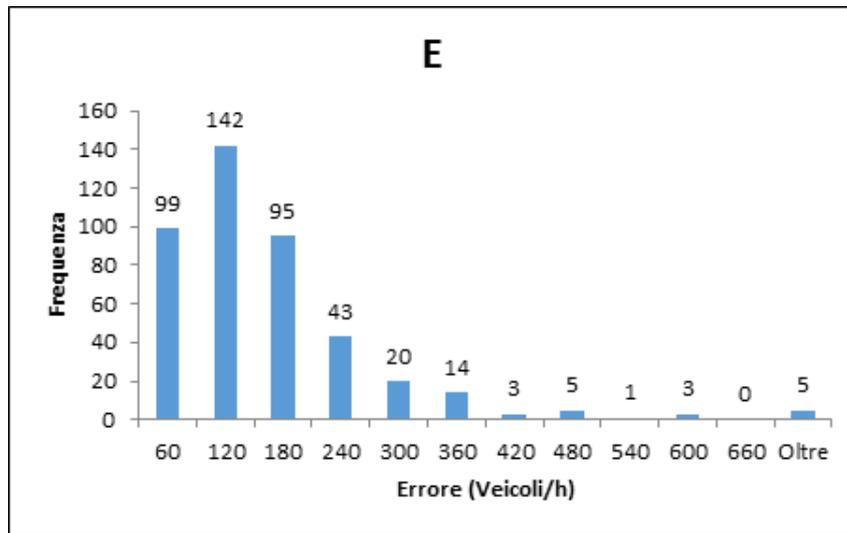
Figura 5.5: Ranking dell'Errore, pesi con tempi, propagazione senza rapporti.



In questo grafico le spire sono ordinate in base all'errore migliore, che è quello calcolato con il metodo del paragrafo precedente.

5.2 Pesì con tempi di viaggio, propagazione senza rapporti di flusso

Figura 5.6: Distribuzione dell'Errore, pesi con tempi, propagazione senza rapporti.



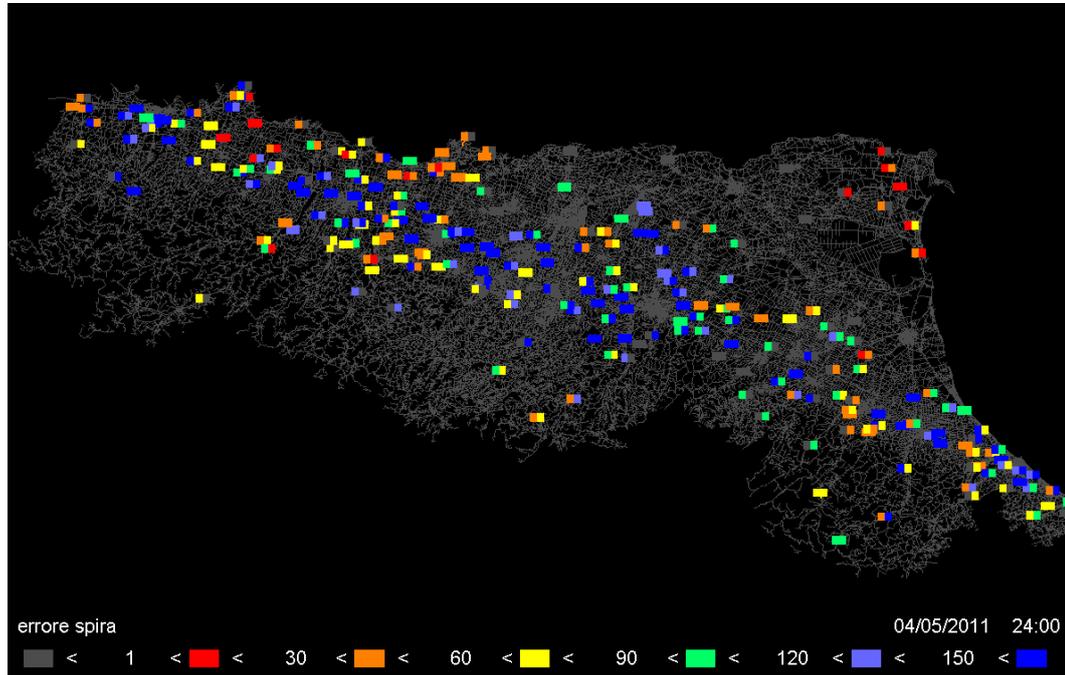
Si osserva subito, dalla scala delle ordinate del primo grafico e dall'istogramma, che l'errore è molto maggiore che nel caso precedente. Il valor medio dell'errore E_m è 138.7 veicoli/h.

È però evidente una distribuzione statistica simile a prima.

Di seguito (Fig. 5.7) è mostrato l'errore sulla mappa.

Questo algoritmo di propagazione non è altrettanto buono del precedente, e anche al di fuori dei centri urbani si vedono aree con errore elevato.

Figura 5.7: Gli errori di ricostruzione dei flussi in rappresentazione geografica, con uno schema di calcolo meno valido.



5.3 Pesi con probabilità di svolta, propagazione con rapporti di flusso

In questo terzo schema, i pesi sono definiti come

$$w_{ji} = p_svolta_{ij} = \frac{n_j^{(out)}}{n_i^{(out)}} \quad (5.7)$$

E la propagazione, esattamente come nel primo metodo:

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p_gps_i}{n_j^{(out)}/p_gps_j} \varphi_j(t) \quad (5.8)$$

Questo metodo dà risultati poco diversi dal primo, poiché il sistema dipende più debolmente dalla scelta dei pesi che dal metodo di propagazione. I due

grafici caratteristici di questo metodo (Figg. 5.8 e 5.9) non mostrano eccessive differenze dal caso suddetto:

Figura 5.8: Ranking dell'Errore con giornata divisa, pesi con probabilità, propagazione con rapporti.

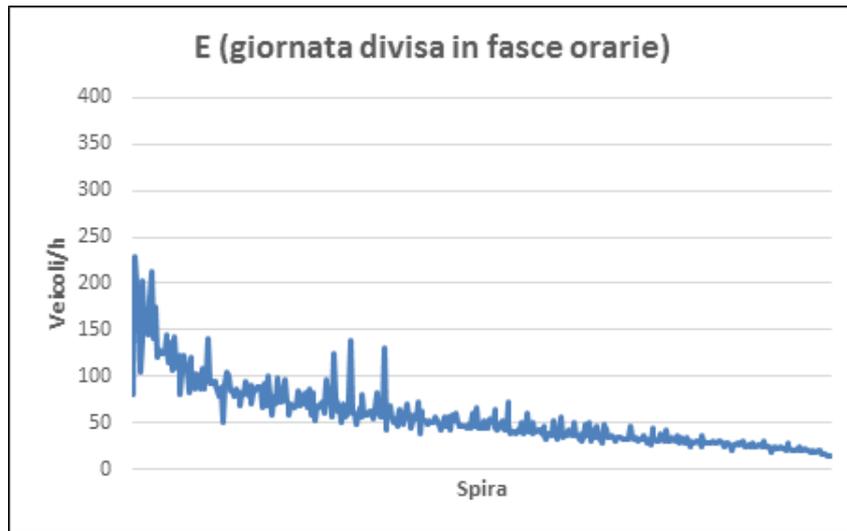
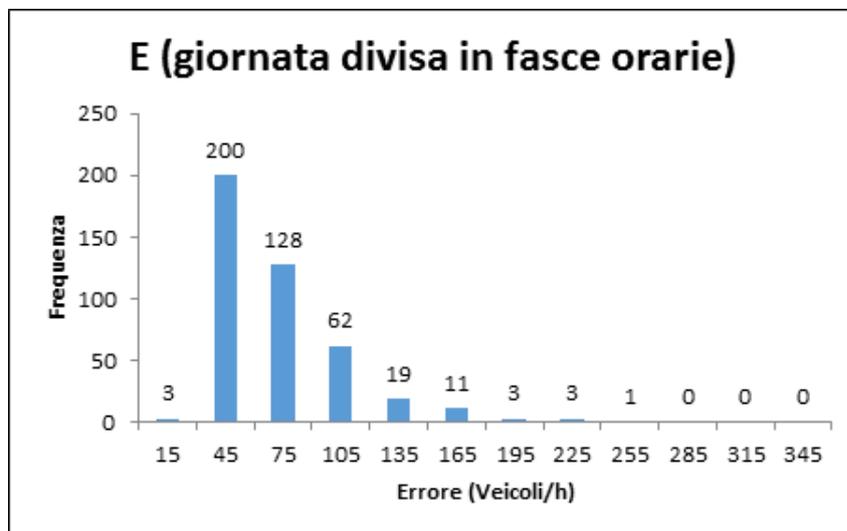


Figura 5.9: Distribuzione dell'Errore con giornata divisa, pesi con probabilità, propagazione con rapporti.



Qui il valor medio dell'errore è di 56.6 veicoli/h. Di nuovo la distribuzione statistica è simile ai casi precedenti.

Si omette la visualizzazione su mappa poiché quasi indistinguibile da quella del primo paragrafo.

5.4 Pesi con probabilità di svolta, propagazione senza rapporti di flusso

Nell'ultimo metodo, di fatto il peggiore in termini di errore, i pesi sono assegnati come

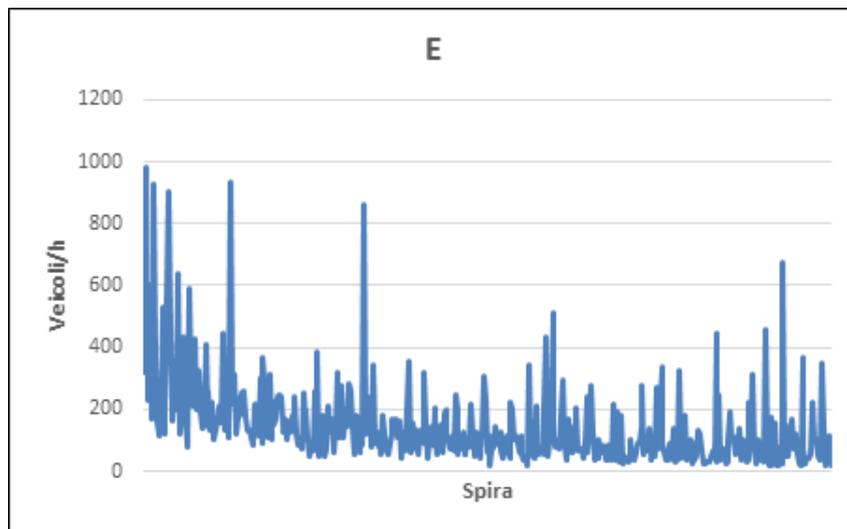
$$w_{ji} = p_svolta_{ij} = \frac{n_j^{(out)}}{n_i^{(out)}} \quad (5.9)$$

E la propagazione avviene con

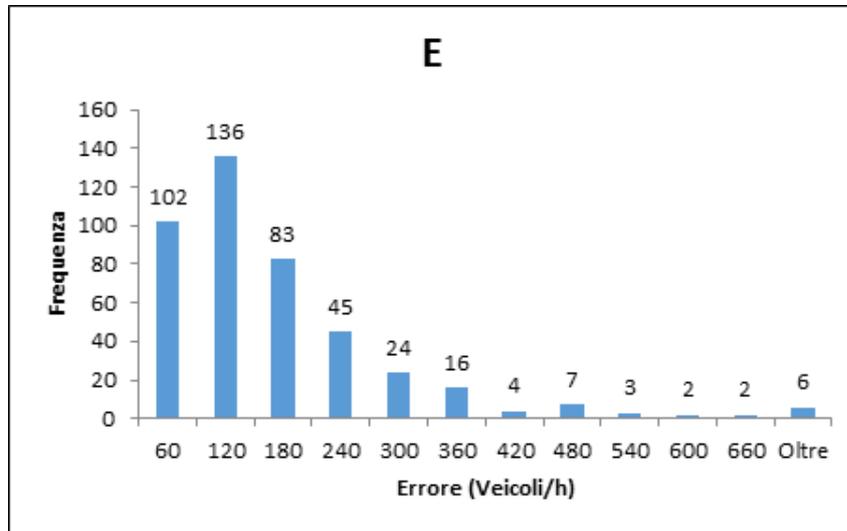
$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t) \quad (5.10)$$

Seguono (Figg. 5.10 e 5.11) i grafici caratteristici:

Figura 5.10: Ranking dell'Errore, pesi con probabilità, propagazione senza rapporti.



Di nuovo le spire sono ordinate secondo l'errore del primo metodo.

Figura 5.11: Distribuzione dell'Errore, pesi con probabilità, propagazione senza rapporti.

Anche qui l'errore è alto, e mediamente vale 147.3 veicoli/h. La distribuzione è quella usuale.

La mappa dell'errore è molto simile a quella del secondo metodo, e verrà omessa.

5.5 Sintesi dei risultati

Tutti i risultati esposti fin ora suggeriscono alcune caratteristiche dei vari schemi di calcolo: innanzitutto si può notare che inserire nella propagazione i rapporti di flusso migliora molto le simulazioni.

Il metodo di propagazione ideale è quindi sempre

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p\text{-gps}_i}{n_j^{(out)}/p\text{-gps}_j} \varphi_j(t) \quad (5.11)$$

In secondo luogo, i risultati mostrano che la bontà del modello dipende molto più fortemente dalla propagazione piuttosto che dalla scelta dei pesi.

Tuttavia, si può vedere che utilizzando gli inversi dei tempi di propagazione

nell'assegnazione dei pesi gli errori sono minori che quando si adottano le probabilità di svolta.

La seguente tabella riassume i risultati numerici delle simulazioni della giornata 4 maggio 2011, mostrando le quattro combinazioni di algoritmi affiancate, dove opportuno, dalla variante con fasce mattina-pomeriggio.

Pesi	Propagazione	Errore (Veicoli/h)	Errore _{MP} (Veicoli/h)
1 / t. propagazione	rapporti flussi	58.9	54.2
1 / t. propagazione	semplice	138.7	-
Prob. svolta	rapporti flussi	61.5	56.6
Prob. svolta	semplice	147.3	-

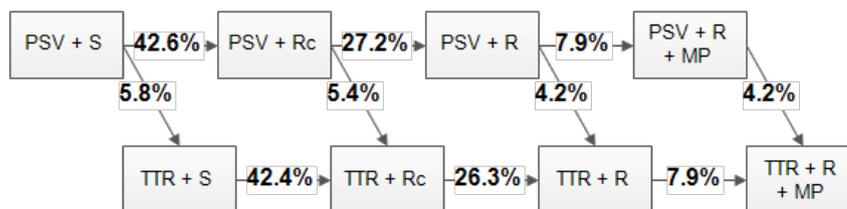
Come già affermato, la variante mattina-pomeriggio produce un miglioramento di circa 7-8%.

Inoltre sono state effettuate diverse simulazioni considerando il parametro di penetrazione pari al 2% per tutte le strade, ottenendo col metodo dei tempi di tempi di propagazione e rapporti di flusso questi risultati:

Penetrazione	Errore (Veicoli/h)	Errore _{MP} (Veicoli/h)
Empirica	58.9	54.2
Fissata a 2%	79.9	-

Volendo schematizzare la validità dei vari algoritmi, si propone il diagramma in Fig. 5.12.

Figura 5.12: Schema della gerachia degli algoritmi.

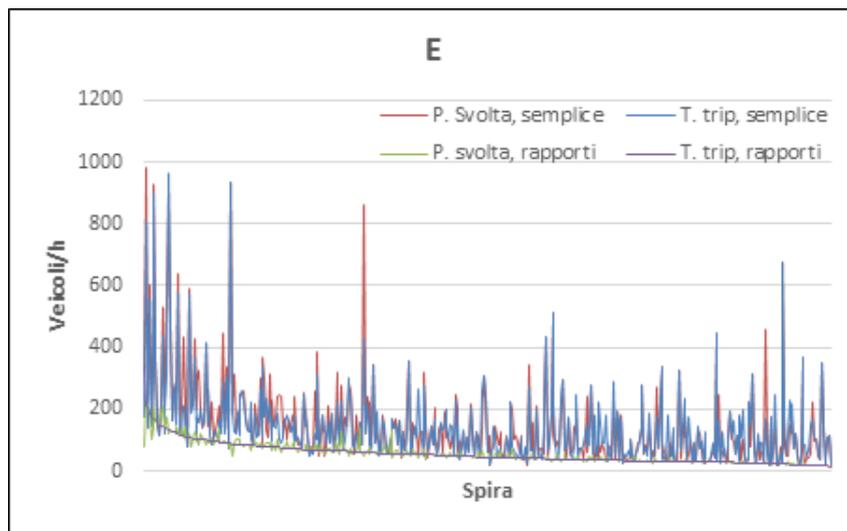


Lo schema mostra le percentuali di miglioramento degli errori di ricostruzione del modello nei passaggi da una combinazione di algoritmi all'altra. Le frecce sono sempre dirette da una combinazione peggiore ad una minore. Si adotta la seguente legenda:

- PSV: probabilità di svolta, TTR: tempi di transito;
- S: propagazione semplice, R: propagazione con rapporti, Rc: rapporti calcolati con penetrazione fissata al 2%;
- MP: divisione della giornata in fasce mattutina e pomeridiana.

Nel grafico di Fig. 5.13 si può osservare un confronto diretto degli errori dei quattro tipi di algoritmo per il modello.

Figura 5.13: Gli errori dei quattro algoritmi a confronto.



Di nuovo si sottolinea che il ranking delle spire è in base all'errore con mattina-pomeriggio del metodo t.trip, rapporti, che è il metodo migliore. Per i due metodi che utilizzano la propagazione con rapporti è mostrato solo l'errore con mattina-pomeriggio.

Sebbene la disposizione ordinata delle spire di questo grafico sia una scelta arbitraria, esso mostra immediatamente un risultato importante: i metodi con

rapporti sono molto più robusti dei metodi semplici alle fluttuazioni spaziali di flusso, cioè nel passare da una spira all'altra.

La ragione di ciò è naturalmente la presenza dei rapporti di flusso nella definizione degli uni e la loro assenza negli altri.

Si evince, inoltre, che i tre metodi inferiori riproducono meglio i flussi solo in pochi punti, ed in misura minima.

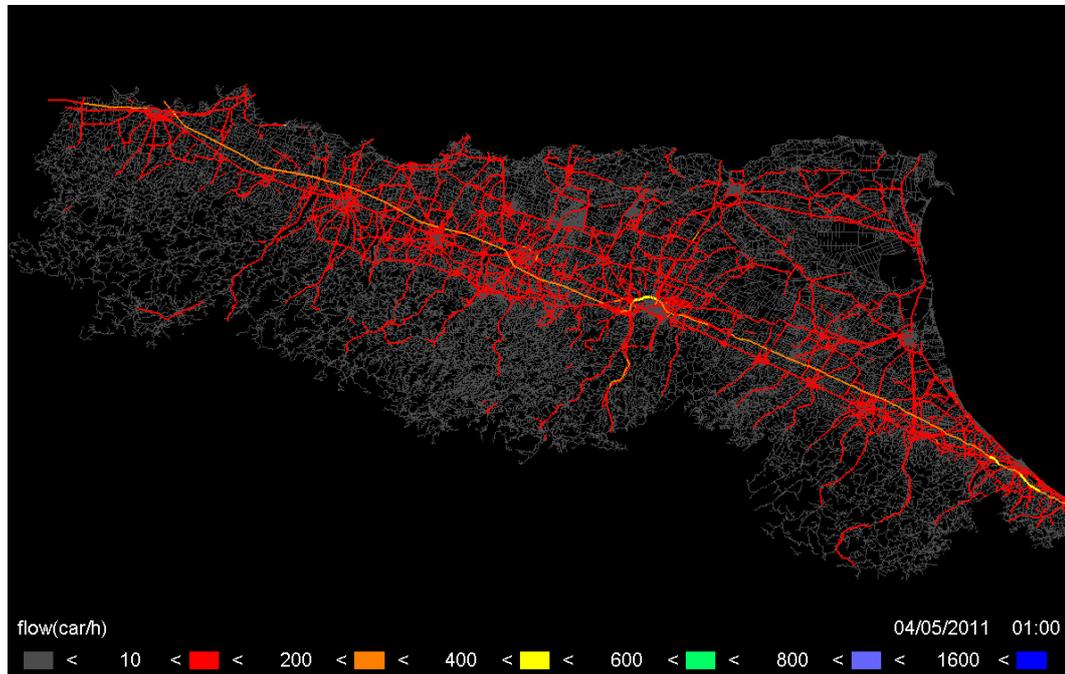
5.6 Calcolo del campo di flusso nella regione

Una volta osservati i risultati dei differenti metodi, è opportuno mostrare il risultato dell'estensione del campo di flusso a partire dai dati di spira a tutta la regione.

Il metodo scelto è quello che produce errore minore, vale a dire il metodo con tempi di transito per il calcolo dei pesi, con rapporti di flusso per le propagazioni e con fasce orarie.

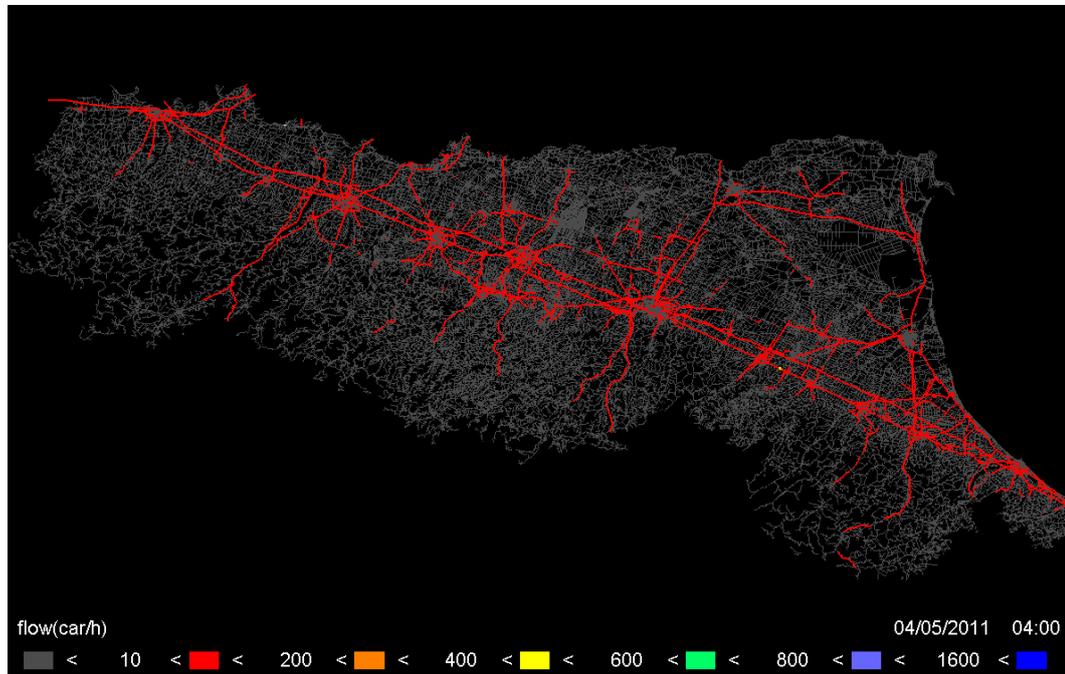
Vengono mostrate di seguito alcune mappe con i flussi ricostruiti dal modello per la giornata del 5 maggio 2011 a diverse ore.

5.6.1 Ore 01:00



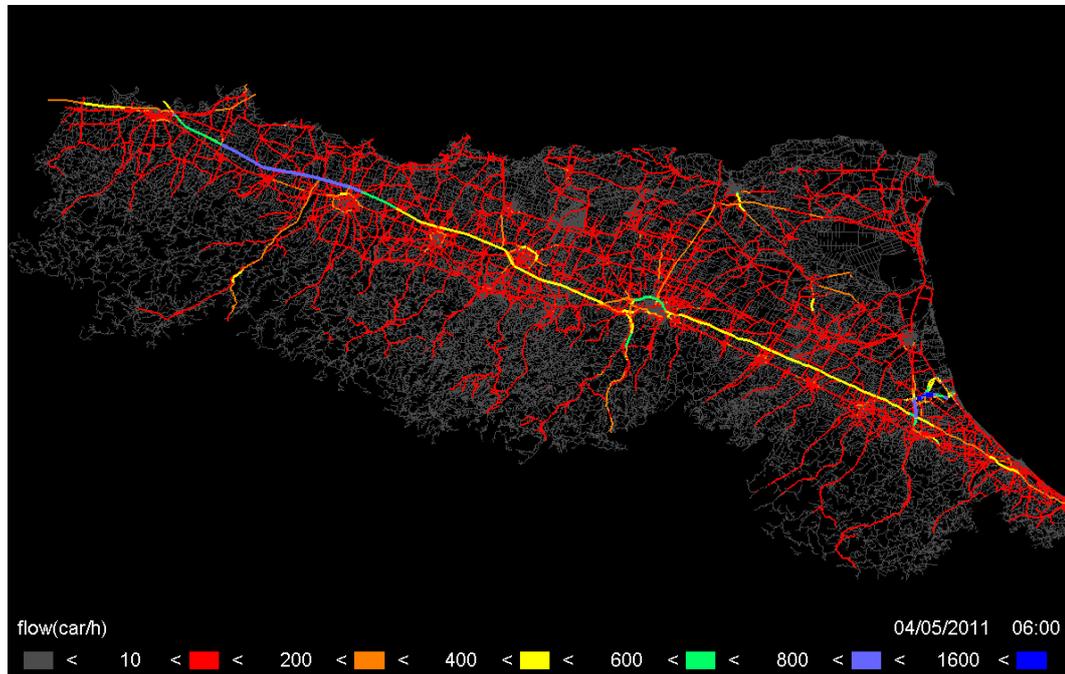
Si osserva come i flussi medi sono molto bassi tranne alcuni flussi notevoli sull'autostrada A1 e soprattutto sulla Tangenziale di Bologna e sulla Riviera verso Rimini. Attorno a queste ore notturne alcune spire hanno mostrato dei flussi relativamente alti, probabilmente dovuti alla chiusura di discoteche o altri locali notturni.

5.6.2 Ore 04:00



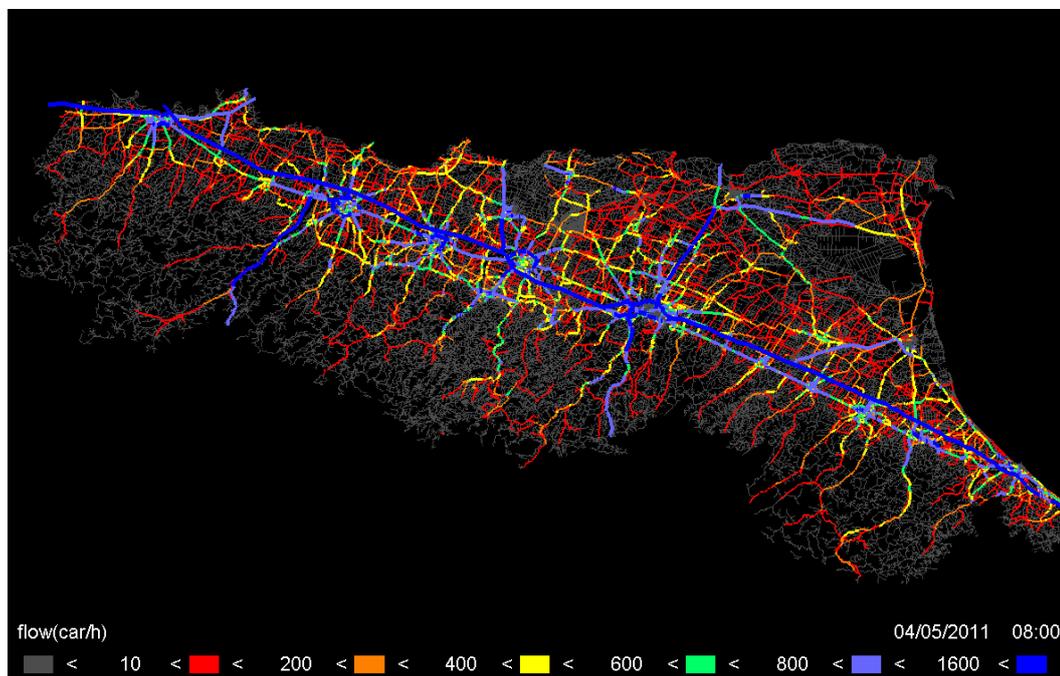
È l'ora in cui il flusso è in assoluto minore su tutto il sistema. Si osservano ovunque meno di 200 veicoli all'ora.

5.6.3 Ore 06:00



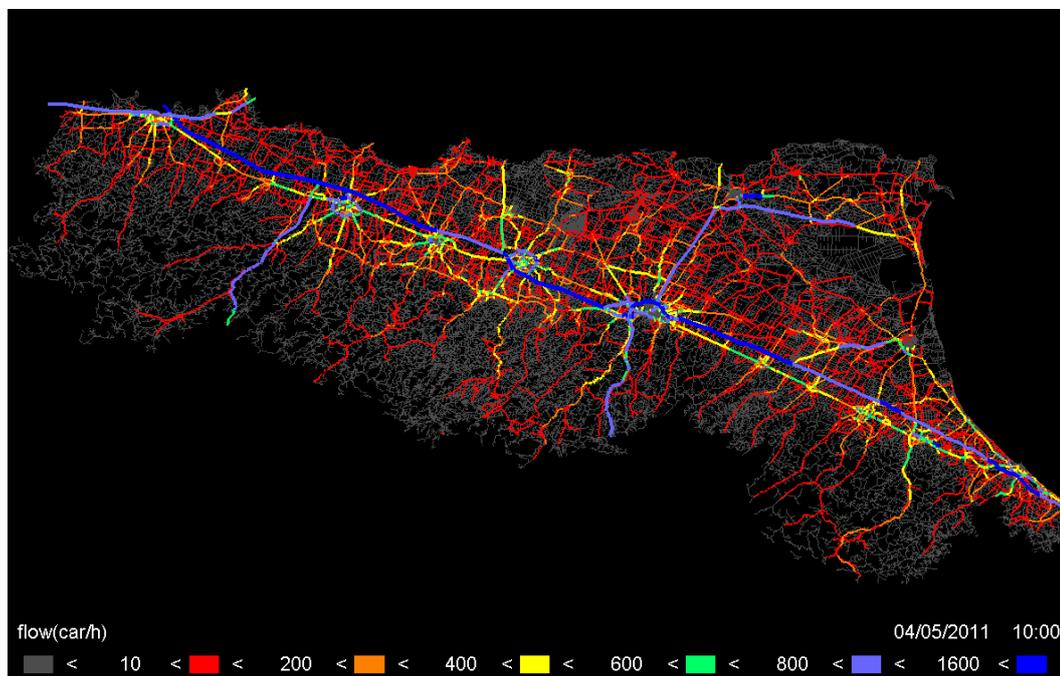
I cittadini iniziano a svegliarsi e le strade iniziano a popolarsi fino a raggiungere il picco mattiniero alle 8:00.

5.6.4 Ore 08:00



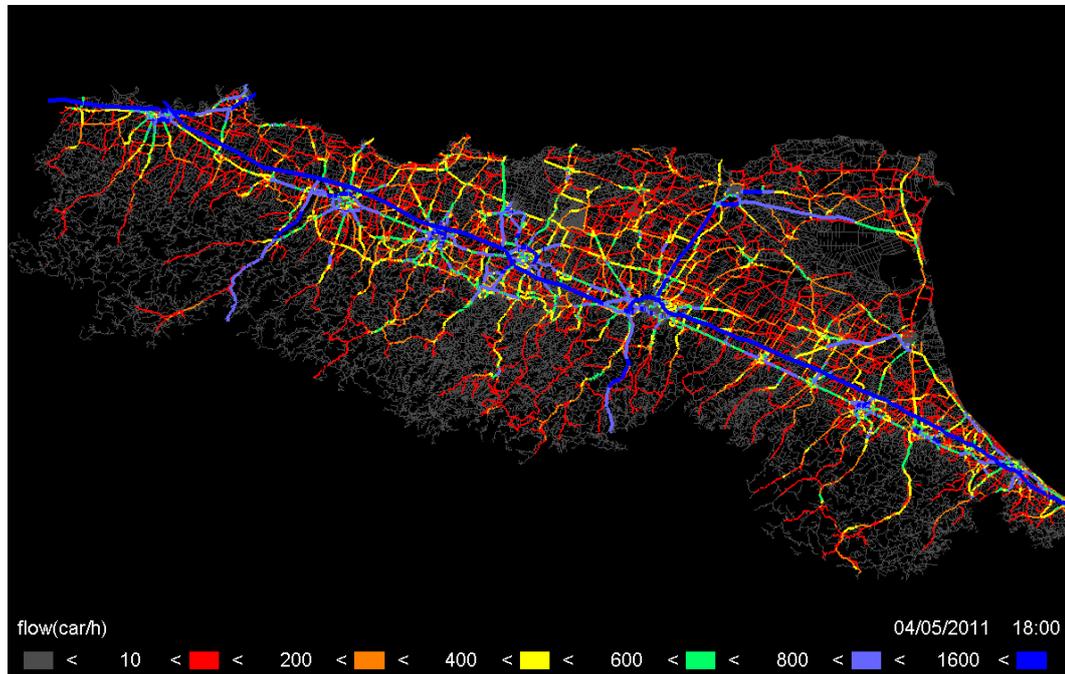
Questo è il principale picco mattutino: i collegamenti principali della rete, come autostrade e superstrade, e Via Emilia, sono molto trafficati, con 1600 e più veicoli/h.

5.6.5 Ore 10:00



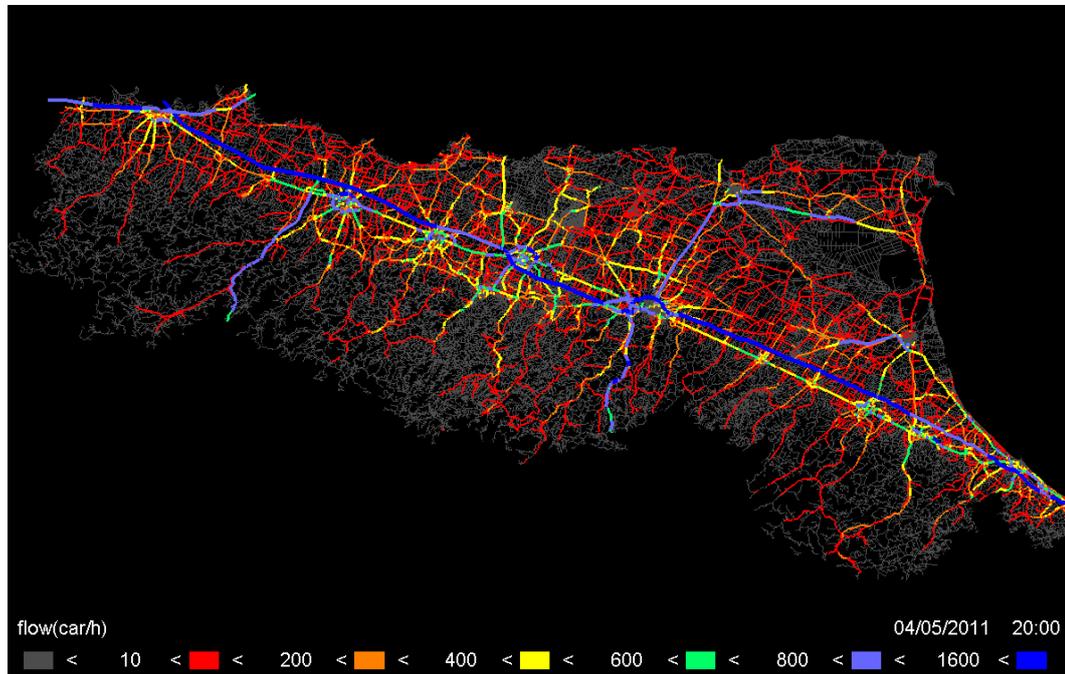
Il flusso si stabilizza: si vedono i collegamenti fondamentali molto trafficati e la situazione rimane all'incirca costante fino alle 17:00, quando il flusso inizia a rialzarsi per raggiungere il massimo alle 18:00. Ci sono piccole variazioni attorno alle 12:00 e alle 14:00 per gli spostamenti dovuti alla pausa pranzo o part time, ma non si riescono ad osservare in figura.

5.6.6 Ore 18:00



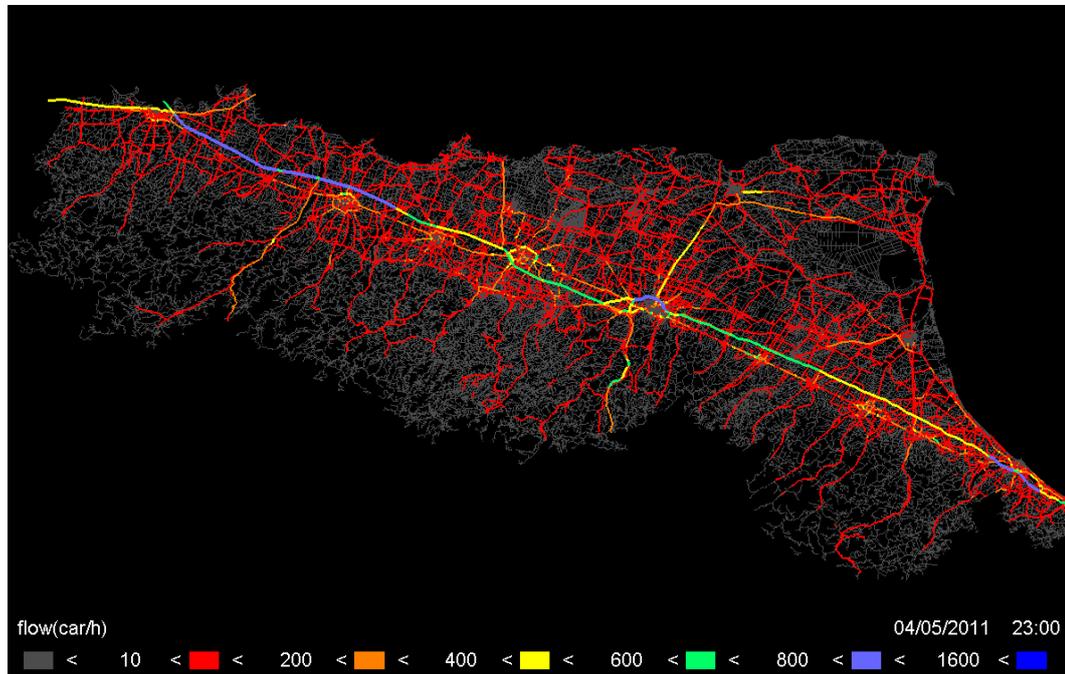
Si osserva il secondo massimo della giornata dove la regione ha mediamente un flusso simile a quello del picco delle 08:00.

5.6.7 Ore 20:00



Si nota che i flussi iniziano a diminuire, trend che continuerà fino a circa le 04:00.

5.6.8 Ore 23:00



A quest'ora si vedono flussi relativamente alti solo nelle arterie principali.

Capitolo 6

Conclusioni

L'intento di questo progetto di tesi era lo sviluppo di un modello matematico e simulativo in grado di ricostruire informazioni di flusso locali su una rete estesa.

La metodologia fondamentale alla base del modello è la conciliazione di due categorie di dati in qualche modo eterogenee; queste due classi di dati sono alla base dell'intero progetto, ed entrambe sono di fatto imprescindibili, ma sono nondimeno affette da limitazioni.

La prima è costituita dai dati provenienti dalle spire; sono misure continue con una buona risoluzione temporale e che si considerano esatte nel contesto del modello, ma sono dati puntuali e nella prospettiva globale hanno una bassa risoluzione spaziale.

La seconda categoria sono i dati dai dispositivi GPS. Essi sono diffusi in tutto il sistema preso in considerazione, ma hanno una bassa risoluzione temporale e sono relativi solo ad un campione di individui non rappresentativo.

Deve risultare evidente che se questi dati fossero perfetti, e non avessero le limitazioni di cui risentono, la loro sinergia sarebbe sufficiente a costruire la conoscenza completa del campo di flusso; e di fatto se le spire fossero onnipresenti, o i dati GPS avessero una capillarità totale, anche una sola di queste due categorie basterebbe ad avere la risoluzione punto per punto ed istante

per istante del campo.

Senza'altro, col passare degli anni, tecniche di misura come le spire induttive e strumenti personali come i GPS assumeranno entità via via più rilevante. Questo porterà ad avere dati più completi, e modelli come quello qui sviluppato diventeranno sempre più accurati e predittivi.

Tuttavia questo lavoro ha mostrato che l'elaborazione e un parziale miglioramento dei limitati dati a disposizione sono possibili, e portano a risultati che si ritengono buoni.

Oltretutto, si è mostrato che il modello è in grado di fornire informazioni spaziali sulla bontà delle proprie ricostruzioni, cioè di mostrare dove produce buoni risultati e dove per varie ragioni è meno affidabile. In un'ottica di miglioramento delle misure, questo fattore è importantissimo, soprattutto per quanto riguarda le spire.

È evidente infatti che se il modello mostra che in particolari zone la ricostruzione non è buona, una pianificazione mirata della disposizione di nuovi strumenti di misura in questi punti critici porterebbe un miglioramento nel modello a costi molto più contenuti di un processo indiscriminato di introduzione di nuove spire.

Il modello è stato sviluppato per la rete stradale della regione Emilia Romagna, ma alcune metodologie applicate hanno naturale estensione anche al resto della rete nazionale.

In primo luogo il concetto di penetrazione comunale (discusso nel paragrafo 2.4.1). Stime di questa grandezza sono note per tutti i comuni della nazione, per cui almeno questa tipologia di penetrazione è già disponibile.

In secondo luogo la diffusione dei GPS. La distribuzione dei dispositivi non è esclusiva della regione Emilia Romagna, ed è una tecnica di indagine assicurativa già adottata in altre zone del paese.

Il discorso è analogo per le spire induttive. Il modello è naturalmente estendibile a ovunque siano disponibili delle reti estese di apparati, come ad

esempio a Roma[4], Milano[11], Torino e quindi in regioni importanti come Lazio, Lombardia e Piemonte.

Bibliografia

- [1] D. J. Tritton, *Physical Fluid Dynamics (second edition)*. Oxford Science Publications (1988).
- [2] M. Treiber & A. Kesting, *Traffic Flow Dynamics*. Springer (2013).
- [3] A. Tizghadam, A. Leon-Garcia, *A Graph Theoretical Approach to Traffic Engineering and Network Control Problem*. 21th International Teletraffic Congress (ITC21), Paris, France, (2009).
- [4] M. G. Messina, G. Valenti, F. Carapellucci, G. Fusco, C. Colombaroni, A. Gemma, G. Ciccarelli, S. Lo Sardo, *Sistema di monitoraggio e previsione della mobilità veicolare per l'integrazione tra la rete della illuminazione pubblica e la rete della mobilità*. http://www.enea.it/it/Ricerca_sviluppo/documenti/ricerca-di-sistema-elettrico/smart-city/rds-330.pdf, p. 93, (2011).
- [5] G. Sabato, *Rilevazione, conteggio e classificazione del traffico mediante spire induttive*. http://www.doingsecurity.it/adon/files/mc2024t6_it_20040116_traffic.pdf (2006).
- [6] S. Rambaldi, *PL10 - Algoritmi di monitoraggio, previsione e controllo*. Report Progetto Pegasus (2012).
- [7] F. A. Haight, *Handbook of the Poisson Distribution*. John Wiley & Sons, (1967).
- [8] *Annuario statistico 2011 ACI*. <http://www.aci.it/> (2011)

-
- [9] R. C. Gonzalez, *Digital Image Processing, third edition*. Prentice Hall (2008).
- [10] B. Piccoli, *Traffic flow on networks: conservation laws models*. <http://hyp2008.umd.edu/files/slides/Piccoli.pdf>, I.A.C.-C.N.R. Roma (2008)
- [11] F. Ferrari, *Flussi di traffico nel comune di Milano: dati rilevati e analisi*. Tesi di Laurea Magistrale in Architettura, Politecnico di Milano, (2011).