

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

Scuola di Scienze
Corso di Laurea Magistrale in Fisica

**RICOSTRUZIONE DI FLUSSI VEICOLARI
SU SCALA REGIONALE: ANALISI DEI
DATI DISPONIBILI**

Relatore:
Prof. Sandro Rambaldi

Presentata da:
Mattia Guidetti

Sessione II
Anno Accademico 2012/2013

Indice

1	Introduzione	5
1.1	Premessa	5
1.2	Contesto	6
1.3	Il problema affrontato	8
1.4	Il modello	10
1.5	Risultati	12
2	Dati a disposizione	14
2.1	La scelta dei dati	14
2.1.1	Caratteristiche giornaliere dei flussi di traffico	15
2.2	Descrizione delle fonti di dati utilizzati	19
2.2.1	Spire	19
2.2.1.1	Installazione	20
2.2.1.2	Funzionamento	21
2.2.1.3	Spire della regione Emilia Romagna	22
2.2.2	GPS	22
2.3	La penetrazione	23
2.3.1	Calcolo senza dati delle spire	23
2.3.2	Calcolo con dati delle spire	24
2.3.3	Estensione con metodo a griglia	25

<i>INDICE</i>	2
3 Algoritmi di Propagazione	30
3.1 Introduzione	30
3.2 Modelli	31
3.2.1 Calcolo dei pesi	31
3.2.2 Propagazione dei flussi	32
3.2.3 Validazione del modello	33
4 Analisi Dati	35
4.1 Elaborazione sulla media delle spire	35
4.2 Correlazione	39
4.2.1 Introduzione	39
4.2.2 Correlazione Triviale	40
4.2.3 Correlazione modulare triviale	41
4.2.4 Correlazione raffinata	46
4.2.5 Correlazione tra influenti e massima	49
4.2.6 Relazioni della correlazione con altre grandezze caratteristiche	53
4.2.7 Gli errori di ricostruzione	56
4.2.7.1 Spira 175, errore basso e distanza di correlazione bassa	57
4.2.7.2 Spira 152, errore alto e distanza di correlazione bassa	59
4.2.7.3 Conclusioni	60
4.2.8 Divisione delle spire in cluster	62
4.3 Parametri di influenza e bontà delle spire	67
4.3.1 La bontà	67
4.3.2 L'influenza spaziale	69
4.3.3 Influenza di flusso	69

<i>INDICE</i>	3
5 Mattina e Pomeriggio, le differenze	70
5.1 Introduzione	70
5.2 Il confronto tra le penetrazioni	71
5.2.1 Divisione in tre fasce	72
5.2.2 Divisione in due fasce	75
5.2.3 Conclusioni	76
5.3 Probabilità di svolta	76
5.4 Validazione della simulazione	83
6 Risultati del modello	88
6.1 introduzione	88
6.2 Risultati delle varie combinazione di pesi e metodi di propagazione	90
6.2.1 Pesi con tempi di viaggio, propagazione con rapporti di flusso	90
6.2.2 Pesi con tempi di viaggio, propagazione senza rapporti di flusso	92
6.2.3 Pesi con probabilità di svolta, propagazione con rapporti di flusso	94
6.2.4 Pesi con probabilità di svolta, propagazione senza rapporti di flusso	95
6.3 Sintesi dei risultati	96
6.4 Calcolo del campo di flusso nella regione	100
6.5 Free Speed	106
7 Conclusioni	107
Bibliografia	109

Abstract

Si mostra l'evoluzione di un modello fisico-matematico atto alla ricostruzione dei flussi nell'intera regione Emilia Romagna partendo da due sorgenti di dati: dati sparsi con bassa risoluzione spaziale senza errore e dati distribuiti in tutta la regione con percentuale rispetto alla totalità di autoveicoli ignota. Si descrive l'elaborazione dei dati e l'evoluzione del modello nella sua storicità fino ad ottenere un buon risultato nella ricostruzione dei flussi. Inoltre si procede ad analizzare la natura dei dati e la natura del problema per ottenere una buona soluzione in grado di descrivere il sistema e per essere facilmente estesa ad altre regioni. Dopo aver definito un metodo di validazione si confrontano svariati risultati di modelli differenti mostrando un'evoluzione nella ricerca di un modello che risolva il problema con tempi computazionali accettabili in modo da ottenere un sistema real-time.

Capitolo 1

Introduzione

1.1 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 dall'altro candidato.

In particolare l'autore di questa tesi è responsabile dell'analisi e comprensione dei dati dal punto di vista dinamico e della fase di validazione dei vari modelli.

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

1.2 Contesto

Il lavoro su cui si basa la presente tesi è un progetto sperimentale che si colloca nell'ambito della mobilità veicolare.

L'idea della gestione e del miglioramento della mobilità è un problema sempre attuale che richiede ancora al giorno d'oggi uno studio approfondito, soprattutto con gli strumenti della fisica dei sistemi complessi, proprio a causa della complessità intrinseca nei sistemi di mobilità e del grande numero di variabili e grandezze che li caratterizzano [1].

Il processo di miglioramento delle reti di viabilità è una questione interessante ed attuale anche per le sue implicazioni pratiche e di conseguenza economiche: facilitare e velocizzare gli spostamenti comporta molti vantaggi. Da un punto di vista sociale spostamenti più agevoli equivalgono per molte persone ad una riduzione dello stress quotidiano; dal punto di vista economico, è evidente che trasporti più veloci conducono a risparmi di denaro e di tempo.

Un risultato molto intuitivo della teoria dei network afferma che aggiungendo collegamenti alla rete la mobilità mediamente migliora. Ciò si traduce, in termini di reti stradali, nel fatto che idealmente la soluzione più semplice per migliorare i collegamenti sarebbe l'aggiunta di strade ad alta percorrenza, come autostrade o superstrade. Spesso tuttavia non è possibile aggiungere strade alla rete, soprattutto per ragioni ecologiche o urbanistiche.

Per queste ragioni il problema si affronta migliorando le infrastrutture ed i collegamenti; la modifica strutturale della rete si concentra quindi sulla qualità anzi che sulla quantità.

Ad esempio, soprattutto nel contesto urbano, la modifica della direzionalità della rete è molto meno dispendiosa dell'aggiunta di strade. Invertire un senso unico o renderlo dove possibile un doppio senso comporta sì conseguenze economiche, ma il costo è limitato.

Nello studio di sistemi complessi come le reti stradali, l'ausilio della simulazione è spesso fondamentale, poiché analisi puramente osservative del sistema non

sono semplici da realizzare in pratica, né tanto meno è semplice soddisfare il fondamentale requisito scientifico della ripetibilità delle osservazioni.

In un sistema complesso come quello della mobilità, gli elementi costituenti hanno scale diverse e seguono una gerarchia: in questo caso gli elementi fondamentali sarebbero i veicoli, i quali si muovono su strade che si ramificano in città, metropoli, regioni, nazioni.

Anche per questo, nella simulazione si può porre la propria prospettiva ad una scala microscopica o macroscopica, ma a seconda del contesto queste scale assumono definizione diversa.

In generale, in un approccio macroscopico al problema si pone l'attenzione su grandezze medie, tralasciando l'analisi dettagliata delle interazioni e dinamiche dei singoli componenti, caratteristiche che invece sono proprio il punto centrale delle osservazioni microscopiche.

Nel caso dello studio della mobilità, ad esempio, un approccio a scala macroscopica potrebbe assumere come unità i centri cittadini o addirittura le regioni, osservando i grandi movimenti di flusso fra queste unità; un approccio microscopico cercherebbe di analizzare le grandezze delle strade o addirittura risolvere la dinamica dei singoli veicoli.

Nella scala microscopica, la scelta di attribuire il ruolo di unità fondamentale ricade nell'altra grande dicotomia di criteri osservativi: quello Lagrangiano e quello Euleriano[1]. Nel primo l'osservazione segue l'individuo, nel secondo l'osservazione è fissa su un punto del sistema; per cui nel primo approccio è spontaneo scegliere il veicolo come unità, e nel secondo osservare l'evoluzione temporale nelle singole strade delle grandezze interessate, quali i flussi.

Per altro, anche nella pratica di una simulazione di un sistema viario, la risoluzione delle singole dinamiche microscopiche può diventare molto pesante da un punto di vista computazionale, mentre il trasporto (in termini di informazione) di grandezze medie come il flusso di una strada resta più agevole.

Il progetto presentato in questa tesi analizza proprio i flussi, assieme ad altre grandezze caratteristiche, delle strade, studiandone l'evoluzione, ed in parte cercando di riprodurla al meglio.

1.3 Il problema affrontato

L'obiettivo di questo lavoro è lo sviluppo di un modello matematico e simulativo che permetta la ricostruzione dell'intero campo di flusso veicolare per una rete stradale, a partire da misure locali e sparse raccolte con alta frequenza.

Nel caso specifico, l'area studiata è l'intera regione Emilia Romagna, ed i dati sono forniti dall'Agenzia della Mobilità della Regione.

La fonte di dati principale di questo modello sono, come verrà poi mostrato più in dettaglio, apparati di misura del flusso su strada, che chiameremo spire; queste misure di flusso costituiscono la base numerica real-time dell'intero modello.

La problematica che il progetto mira a risolvere, quella della ricostruzione di informazioni di flusso, è fondamentale negli studi sulla viabilità delle reti stradali. Infatti, la conoscenza di grandezze caratteristiche delle strade come, per citarne solo alcune, flusso medio, massimo, velocità e densità di veicoli, costituisce un'informazione preziosa per lo studio e quindi per l'eventuale gestione del traffico, poiché queste grandezze danno indicazioni sulle zone critiche della rete o su quelle poco interessate dal traffico.

Il problema nasce dal fatto che misure dirette e diffuse sono difficili da attuare, e spesso si dispone, come nel caso di questo progetto, di misure sparse, discontinue nello spazio e nel tempo.

Nello studio presente, i dati a disposizione provengono, come accennato, da spire induttive in grado di rilevare i veicoli che transitano sulla strada. Esse sono state installate dall'ente regionale per la mobilità sulle strade principali della rete.

Questi dispositivi, di forma solenoidale e denominati spire, sono posizionati sotto il manto stradale. Questi apparati hanno la possibilità, oltre a conteggiare i veicoli, di valutarne la velocità e di classificarli, basandosi sul tempo di transito, in diverse categorie veicolari:

- motoveicoli
- auto monovolume
- auto monovolume con rimorchio
- furgoncini/camioncini

- camion medi
- camion grandi
- autotreni
- autoarticolati
- autobus

Il funzionamento delle spire verrà spiegato più in dettaglio nel “Capitolo 2”.

È chiaro che una volta ottenuti i dati GPS, la ricostruzione deve passare dalla progettazione e implementazione di un modello matematico, ma nondimeno è importante la fase di studio ed elaborazione dei dati.

Questi dati provenienti dalle spire possono essere considerati privi d’incertezza, poiché tutte le altre sorgenti d’errore del modello sono molto più critiche.

La conoscenza storica del sistema fa da guida nell’estensione del campo di flusso e 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 (dato calcolato confrontando le automobili dotate di sensore GPS rispetto a quelle immatricolate in tutt’Italia).

Ovviamente, anche questa seconda fonte di dati ha necessitato di analisi ed elaborazioni prima di poter essere utilizzata.

Da questi dati è possibile risalire al campo di flusso che purtroppo risulta molto impreciso per la disomogeneità della penetrazione statistica dei dispositivi di raccolta. La determinazione di questo importante parametro è una parte essenziale di questo lavoro, e come si vedrà si sono sperimentati diverse metodologie per arrivarci.

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. È proprio questo moto, quasi diffusivo, con una scala spaziale di una decina di chilometri che fa variare lentamente la penetrazione nello spazio.

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. Infatti, 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'illustrazione delle varie analisi condotte sui dati raccolti sarà lo scopo principale di questa trattazione, e si mostrerà che questi studi hanno guidato la buona riuscita del modello.

1.4 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 "Figura1.1".

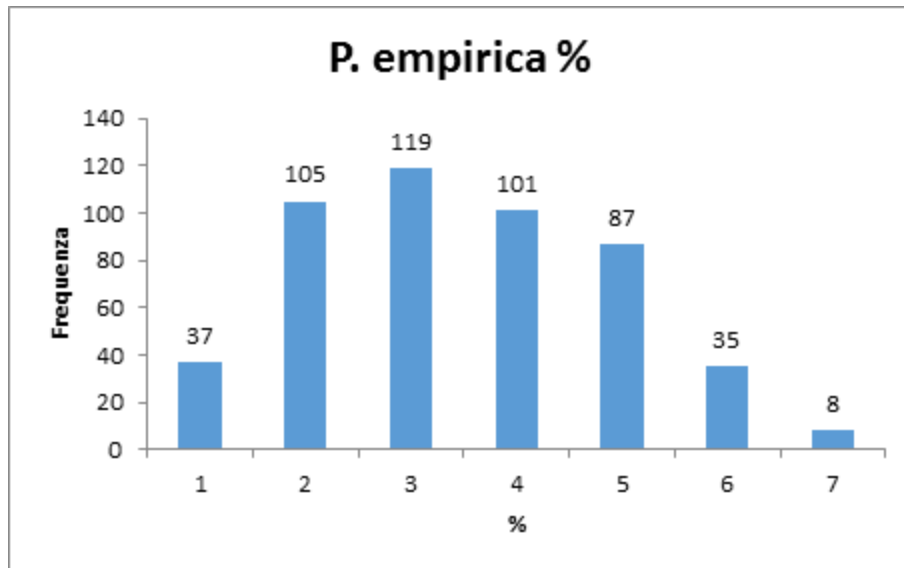


Figura 1.1: questo grafico rappresenta le penetrazioni di ogni spira calcolate come il rapporto tra le auto totali dotate di un GPS che passavano sulla strada sulle auto totali rilevate dalle spire.

Si è 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 fase si è propagata l'informazione del flusso dalle strade con le spire verso tutte le altre utilizzando i risultati della prima fase.

Naturalmente nessun modello è perfetto e produce una ricostruzione totalmente accurata. Non ci si può esimere dalla valutazione della correttezza di un modello, cioè da una fase di validazione.

In questo progetto si è valutata la bontà delle ricostruzioni con un processo di validazione che si articola nei seguenti passaggi:

- 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".

1.5 Risultati

Utilizzando le metodologie sopracitate si è costruito un modello che riesce a rappresentare i flussi nell'intera regione nelle strade considerate con una buona approssimazione. Si sono studiate diverse tipologie di modelli e si sono confrontati i risultati ottenuti cercando il modello che permettesse di ottenere il miglior parametro della validazione compatibilmente con la velocità di calcolo per rendere il modello real time, cioè in grado di fornire i risultati istante per istante.

Gli studi che sono stati effettuati per ottenere questo risultato sono stati i seguenti:

- Pulitura dei dati
- Calcolo della penetrazione
- Elaborazione dei dati per ottenere i parametri caratteristici del modello
- Confronto dei differenti modelli
- Suddivisione della giornata in due fasce dividendo mattina e pomeriggio dove necessario

Attraverso questi studi e differenti processi di modifica del modello si è riusciti a ridurre l'errore commesso nel stimare le spire, come spiegato nella validazione, e quindi ci si ritiene soddisfatti.

Capitolo 2

Dati a disposizione

2.1 La scelta dei dati

Il modello che ci si propone di realizzare si sviluppa su un set limitato di dati: i dati a disposizione del gruppo di lavoro sono quelli da GPS del mese di maggio 2011 ai quali sono stati affiancati quelli delle spire nello stesso periodo.

La prima analisi fatta è stata sulla coerenza temporale dei dati durante il mese, cioè si sono cercati comportamenti diversi nei diversi giorni della settimana. Si è evidenziato che i flussi hanno un comportamento molto differente nelle giornate feriali rispetto a quelle festive. Come esempio in figura si mostrano i valori di flusso medio diurno durante il mese.

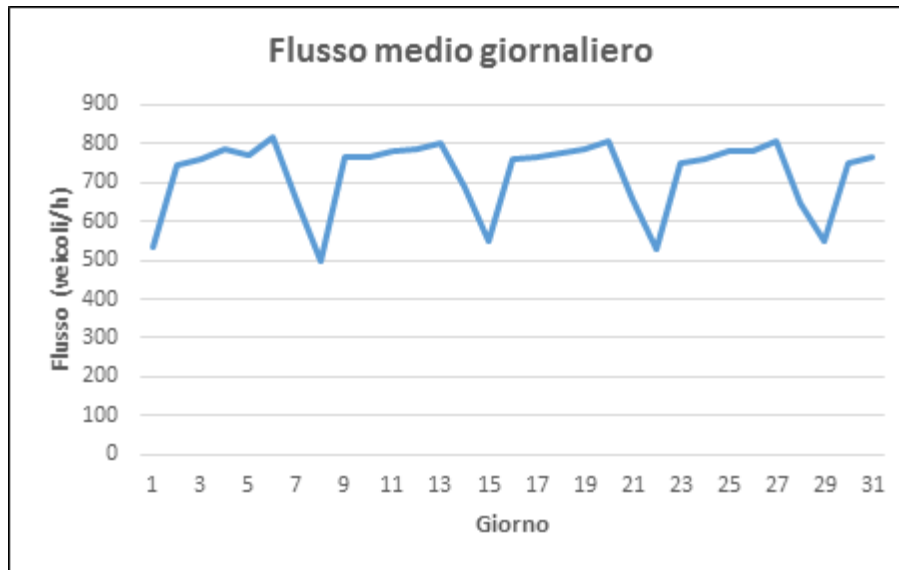


Figura 2.1: è rappresentato l'andamento medio giornaliero della spira 1129 nel mese di maggio del 2011

La figura è molto rappresentativa e mostra un andamento dei flussi con valori più bassi la domenica ed il sabato e con valori più alti nei giorni feriali con una lenta crescita dal lunedì al venerdì.

L'interesse di questo lavoro è quello di creare un modello che ricostruisca i flussi nella regione in maniera generale, astraendosi da una scelta particolare di giorno. Si è quindi deciso di studiare inizialmente le giornate feriali, 20 giorni su 31, che si è osservato essere abbastanza coerenti fra loro e indicative di un comportamento generale.

Ci si propone di trovare una relazione tra i dati GPS e quelli delle spire, e una volta ottenuta questa relazione la metodologia utilizzata sarà estendibile anche alle giornate festive. Si è deciso di partire dai giorni feriali poiché sono in numero maggiore e quindi forniscono una statistica più accurata, inoltre i giorni festivi sono meno coerenti in quanto più soggetti a fluttuazioni di comportamento dovute anche a fenomeni metereologici [2].

2.1.1 Caratteristiche giornaliere dei flussi di traffico

I dati che sono stati analizzati per questo modello sono principalmente i dati di flusso giornaliero di autoveicoli su una determinata strada; sono quindi un tipo

di dati caratteristici di un metodo euleriano, in cui la posizione è fissata e si studia la variazione temporale del sistema in quel determinato punto.

I valori del flusso sono condizionati dalla componente giornaliera. La maggior parte dei dati studiati sono simili e seguono l'andamento qui mostrato:



Figura 2.2: è mostrata la media dei flussi delle 100 spire più popolate della giornata del 4/5/2011.

Nella seguente figura sono invece graficati gli andamenti delle prime 30 spire più popolate nella stessa giornata.

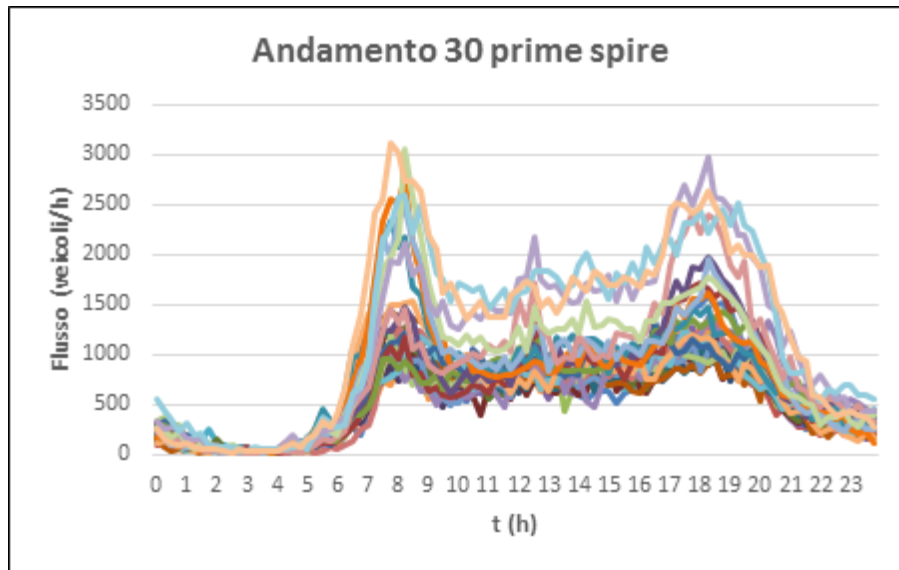


Figura 2.3: sono mostrate le prime 30 spire ordinate per flusso medio della stessa giornata

Da questi grafici è possibile notare che l'andamento del flusso giornaliero può essere diviso in 4 zone:

- 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 6 fino ad arrivare al suo massimo attorno alle 8 per poi calare fino alle 9;
- poi una zona più o meno contenuta e più o meno costante durante la giornata, con un picco attorno all'ora di pranzo, fino ad arrivare alle 16 dove il flusso aumenta nuovamente per crescere fino alle 18 e raggiungere il massimo;
- infine la discesa fino a verso le 20.

Questo andamento è dovuto al fatto che la mattina la città è molto trafficata a causa della tendenza della gente a raggiungere il posto di lavoro e la pendenza dell'aumento è molto alta e ben localizzata, mostrando quanto gli spostamenti

siano sincronizzati su tutta la regione; dopo una zona in cui il flusso rimane molto alto, si ha una discesa quando gli individui raggiungono il proprio posto di lavoro. Durante il giorno comunque il flusso di autoveicoli rimane alto, poiché molte persone continuano a muoversi durante le ore diurne.

In alcuni casi si osservano due piccoli picchi tra le 13:00 e le 14:00, che indicano che una parte delle persone si sposta per pranzare e per tornare o a casa o al lavoro. Il flusso si mantiene all'incirca costante fino a che gli individui verso le 16:00 - 17:00 si rimettono in viaggio per tornare a casa. La coda di questo frangente è molto più lunga rispetto all'andamento mattutino poiché durante la serata le persone tendono a svolgere diverse attività che li mantengono in movimento fino a che, attorno alle 22:00, una buona parte degli individui è rincasata; dopo di che il flusso si mantiene basso fino alla mattina successiva.

La statistica di questo sistema è dominata dal picco mattutino, 7:00 - 9:00, e dal picco pomeridiano, 17:00 - 19:00, durante la quale la maggior parte dei veicoli sono in movimento. Poiché ci si aspettano possibili differenze nelle rispettive statistiche si è deciso di dividere la giornata in due parti, dalle 2:00 alle 14:00 e dalle 14:00 alle 2:00 del giorno dopo, per separare questi due picchi.

Per evidenziare la differenza nelle proprietà statistiche in “Figura 2.4” si mostra l'andamento orario dei flussi nei due versi di percorrenza in una strada ad alto traffico. Si osserva che i due picchi non assumono valori di flusso uguali, andando a sottolineare come alcune strade siano scelte preferibilmente la mattina ed altre la sera; ad esempio una strada che porta da un piccolo borgo a una grande città verrà imboccata in un senso di marcia alla mattina, quando gli abitanti del piccolo borgo vanno a lavorare in città, mentre sarà presa nel senso opposto alla sera. Osservando il grafico si può notare come le altezze dei picchi si invertano.

Altre strade non mostrano questa distinzione.

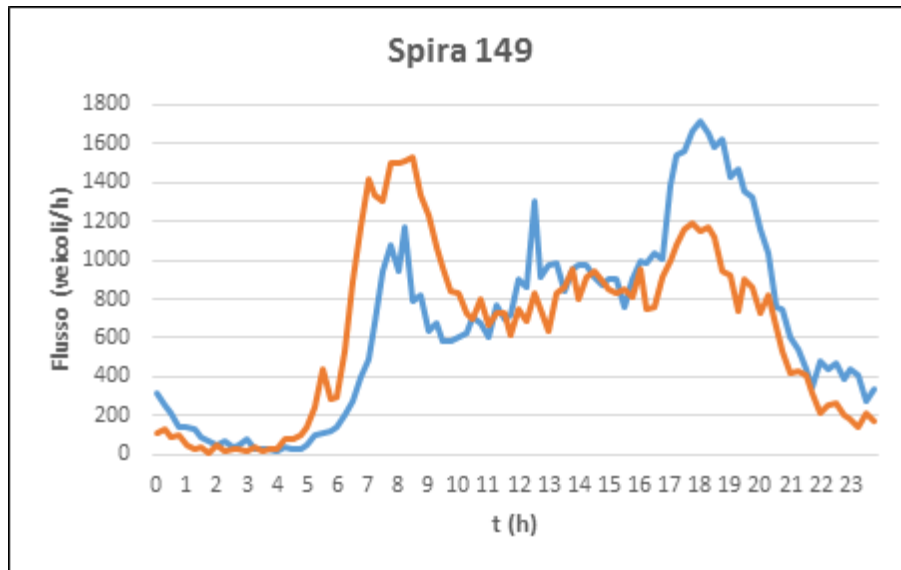


Figura 2.4: è mostrato il flusso medio giornaliero della spira 149 nei feriali del mese di maggio 2011

2.2 Descrizione delle fonti di dati utilizzati

2.2.1 Spire

Con il termine “spira” si intende una coppia di spire induttive per direzione di marcia posizionate sotto al manto stradale come mostrato in figura

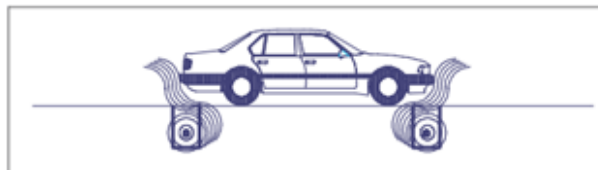


Figura 2.5:

Queste 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[3].

2.2.1.1 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 "Figura 2.6".

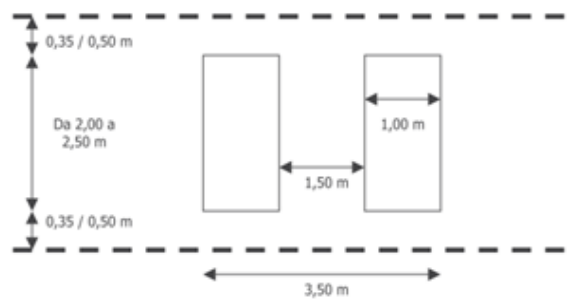


Figura 2.6: Descrizione schematica dell'apparato

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.



Figura 2.7: Foto dell'inserimento del cavo nel cemento

Posizionato come in “Figura 2.7” 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.

2.2.1.2 Funzionamento

I due set di cavi arrotolati 50 volte nel terreno sono percorsi da correnti alternate le quali inducono un campo magnetico, anch'esso di conseguenza alternato, come si nota in “Figura 2.8”. 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.

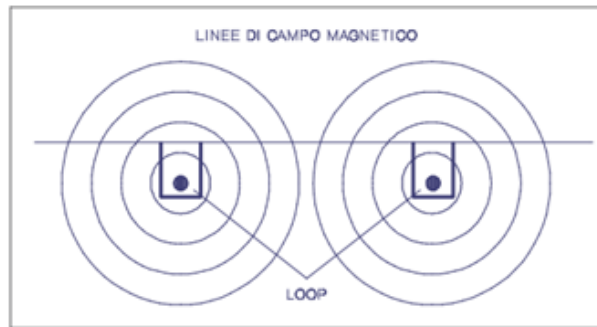


Figura 2.8: Foto dell'inserimento del cavo nel cemento

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 analizza le variazioni lente delle oscillazioni, dovute ad esempio da cambiamenti di temperatura, e quindi è in grado di distinguere il passaggio di veicoli da eventi climatici.

2.2.1.3 Spire della regione Emilia Romagna

Nella regione Emilia Romagna sono state inserite 278 postazioni funzionanti secondo i principi esposti precedentemente. È importante considerare che ogni coppia di spire fornisce due set di dati differenti: un set per ogni verso di percorrenza della strada.

Le spire sono state posizionate nelle strade più importanti della regione, escludendo i tratti autostradali; la loro distribuzione è quindi disomogenea nello spazio, ma essendo state poste nelle zone più critiche per la mobilità, hanno una concentrazione maggiore nelle zone più significative.

I risultati delle misure di queste spire sono stati forniti come flussi di autoveicolo transitati sul tratto di strada ogni 15 minuti. Si è osservato che i dati erano incompleti e con diverse patologie da correggere, per un'analisi più approfondita di questo si osservi la tesi del mio collega. I dati da spira in questo lavoro di tesi verranno considerati corretti e le varie incertezze rimosse.

2.2.2 GPS

Il set di dati è raccolto da un dispositivo (GPS/GSM) chiamato black box che viene distribuito da diverse compagnie di assicurazioni a clienti interessati. Il

problema legato ad una distribuzione dei dispositivi in tal modo ha portato ad una distribuzione non omogenea degli apparati.

Dopo diversi studi [5], che sono esplicitati nel lavoro di tesi del mio collega si è scelto di escludere dalle analisi le strade in cui fosse presente un basso flusso perché la ricostruzione a partire dalla penetrazione richiede che ci sia un certo numero di dati per essere efficiente. L'errore, se si decidesse di ricostruire ogni strada, aumenterebbe in grande misura andando ad inficiare in maniera molto evidente sulla ricostruzione dei flussi sull'intera regione.

2.3 La penetrazione

Per poter utilizzare i dati da GPS con efficacia è necessario conoscere con precisione la frazione delle auto dotate di strumentazione GPS rispetto alla totalità delle auto; questo valore è definito penetrazione. Questo parametro è di grande interesse per lo studio del sistema perché permette di ricavare informazioni sul flusso dei veicoli a partire dai dati sparsi del sistema, riuscendo così a confrontare i flussi con i limiti orari delle strade ottenendo informazioni sul traffico e sulla congestione delle strade. Il problema del calcolo di questo valore è che gli strumenti GPS non sono stati distribuiti in maniera uniforme a tutta la popolazione, infatti si è riscontrato che l'acquisto di questi strumenti è stato maggiore da parte di alcune fasce di età, di reddito e comune. Al momento non è possibile indagare ulteriormente su questo poiché le leggi sulla privacy non permettono di attribuire ad un individuo caratteristiche quali età, tipo di auto e fasce di reddito.

I valori di penetrazione sono indispensabili per lo sviluppo del modello e quindi sono stati calcolati con accuratezza e con differenti metodologie per poter classificare zone e poi poter estendere questi valori anche a regioni in cui non è possibile confrontare i dati con i valori delle spire.

2.3.1 Calcolo senza dati delle spire

Per calcolare i valori di penetrazione senza utilizzare i dati da spira si sono seguiti i seguenti passaggi:

- Valutazione del numero di auto con cui confrontare i dati GPS
- Attribuzione di un comune di appartenenza alle auto dotate di GPS

- Confronto tra i diversi valori 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 [6].

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 questi 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.

La penetrazione calcolata in questo modo verrà chiamata comunale.

Questa modalità di calcolo del parametro di penetrazione è stata implementata traendola da progetti affini a quello qui presentato che non avevano a disposizione dati da spire, ma necessitavano della conoscenza di questa grandezza.

2.3.2 Calcolo con dati delle 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 rapportando i due valori è quella esatta.

Come primo passo si calcolato questo valore nelle strade con spire, 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 nel "Capitolo 3". Questo valore di penetrazione verrà chiamata penetrazione empirica.

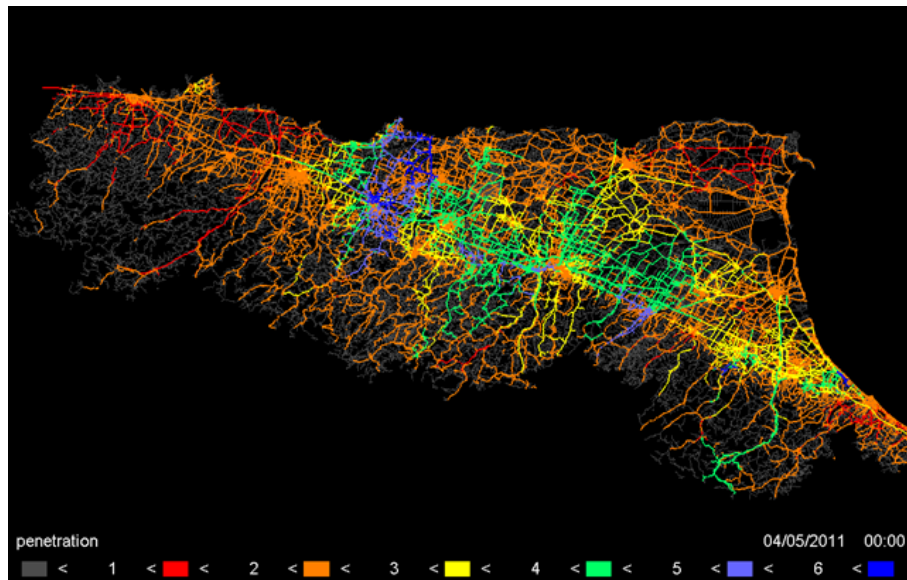


Figura 2.9: Risultato dell'estensione della penetrazione esatta, Penetrazione Empirica

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

2.3.3 Estensione con metodo a griglia

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 spira all'interno di quel comune e le penetrazioni dei comuni adiacenti, era un'approssimazione probabilmente non corretta.

È stata creata una griglia quadrata immaginaria ed è stata posizionata sul centro della spira e si è poi applicata una media pesata dei diversi punti di questa griglia 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 che tenesse conto dei comuni adiacenti.

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 mediata.

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.

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.

quadrato		cerchio	
Raggio Km	P griglia - P esatta	Raggio Km	P griglia - P esatta
8	388.4	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	386.0	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	386.0
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

Tabella 2.1: si osserva la somma delle differenze tra gli errori commessi utilizzando la penetrazione a griglia rispetto a quella esatta, per dimensione della matrice, utilizzando una griglia 11x11 in entrambe le simulazioni

Si può notare come la distanza tra il valore di penetrazione esatta e quella a griglia non cambi in maniera molto forte rispetto alla dimensione del raggio del cerchio o il lato del quadrato considerato finché questo valore rimane più piccolo della grandezza dei comuni si ottengono risultati simili. Si è scelto di utilizzare come valore di raggio 10.25 km con una griglia a forma quadrata.

Questo metodo per calcolare valori di penetrazione è stato successivamente ap-

plicato 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, simile 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[4].

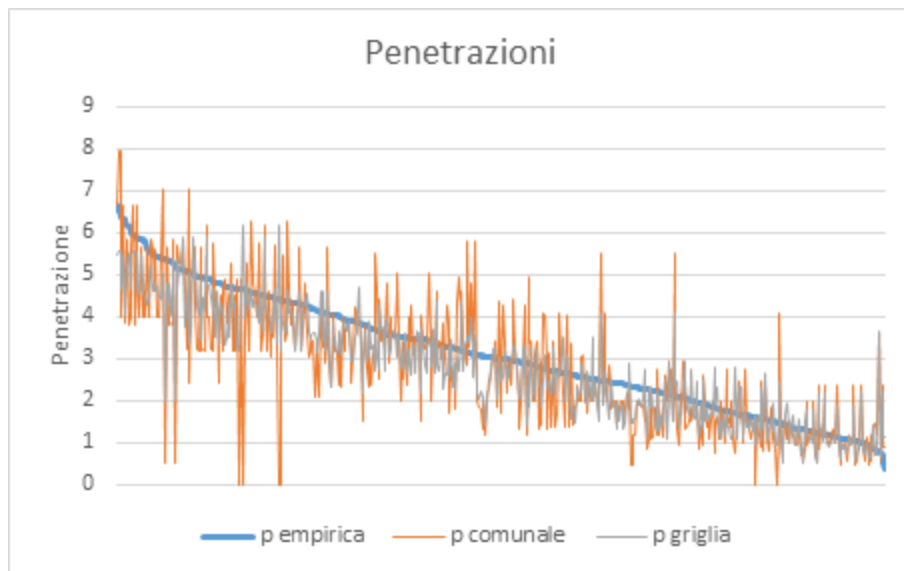


Figura 2.10: Si osservano in figura i confronti tra i differenti metodi di calcolo di penetrazione ordinati secondo la penetrazione empirica

Nel “Figura 2.10” è mostrato un confronto fra i tre tipi di penetrazione calcolati, ordinando le spire in base al valore di penetrazione empirica.

Nonostante la penetrazione calcolata con la griglia 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 griglia si ottengono risultati peggiori rispetto a quelli ottenuti con la penetrazione empirica, che sfrutta i dati GPS, ma migliori che utilizzando quella comunale;

- 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

È importante ricordare i dati a disposizione per la realizzazione del modello, in quanto in questo lavoro di tesi l'obiettivo è quello di ricostituire i flussi su tutta la rete della regione, utilizzando al meglio le informazioni disponibili.

I dati disponibili sono quelli sulle auto e quelli sulla cartografia per quanto riguarda le informazioni delle automobili:

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

Invece per la cartografia si possiede una “cartografia” digitale della rete regionale 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.

I modelli sviluppati non hanno la pretesa di simulare perfettamente e interamente la rete della regione Emilia Romagna, infatti si è fatta una selezione di

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

3.2 Modelli

Il modello come già precedentemente citato è composto da tre passaggi principali, la propagazione dell'informazione e l'estensione di questa in tutto il sistema più un calcolo della correttezza di questo chiamata validazione.

- 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;
- 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;
- 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.

3.2.1 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.

Quelli utilizzati sono i seguenti:

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

$$w_{ji} = \frac{1}{t_trip_{ij}} \quad (3.1)$$

- 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.

$$w_{ji} = p_svolta_{ij} = \frac{n_j^{(out)}}{n_i^{(out)}} \quad (3.2)$$

Dove i è la strada in cui si arriva e j è la sorgente dell'informazione, w è il valore del peso, t_trip è il tempo necessario per andare dalla strada j a quella i e p_svolta_{ij} è la probabilità della svolta

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 (p_gps_i) per ogni strada come:

$$p_gps_i = \sum_j w_{ji} p_gps_j \quad (3.3)$$

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

3.2.2 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:

- 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:

$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t) \quad (3.4)$$

Nell'ultimo metodo sviluppato 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

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p_gps_i}{n_j^{(out)}/p_gps_j} \varphi_j(t) \quad (3.5)$$

Dove $\varphi_i(t)$ è il flusso nella strada i , $n_i^{(out)}$ è il numero totale delle auto uscenti dalla strada i .

Questa tipologia di propagazione ricalca l'idea che il sistema più che essere dominato dalla propagazione dei flussi è dominato da una sorta di trend comune esteso a tutto il sistema, cioè muoversi in maniera isocrona, cioè che il ritardo temporale dell'informazione propagata tra la spira ed il punto di arrivo non influisce sul risultato poiché il sistema tende a muoversi in maniera coerente. Per avere un'analisi più accurata del modello è consigliabile studiare la tesi del mio collega.

3.2.3 Validazione del modello

Precedentemente si è parlato della combinazione di parametri che permette di ottenere l'errore minore ed il metodo con cui si è calcolato questo errore è il sistema di validazione, che permette di ottenere un parametro di bontà nel modello.

Nel presente lavoro, la validazione è basata sull'attribuzione di un valore di errore massimo 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. È importante considerare che la spira nella direzione opposta non viene considerata nel sistema.

Per ottenere questi errori, dapprima si simula tutto il periodo temporale di interesse (che qui si assume sia una giornata), e alla fine della simulazione, per ogni spira j e ad ogni orario t , si calcola il rapporto seguente:

$$\epsilon_i(t) = |\varphi_i^{(misur)}(t) - \varphi_i^{(calcol)}(t)| \quad (3.6)$$

L'errore finale è quindi la somma su tutto il periodo e su tutte le spire di questi errori parziali, indipendentemente dalla tipologia utilizzata. Formalmente:

$$E = \sum_i \int_{t_0}^T \epsilon_i(t) dt \quad (3.7)$$

È importante ricordare che l'errore calcolato rappresenta l'errore che si commetterebbe nel ricostruire il flusso della spira se essa non fosse presente nel sistema, ed è quindi un upper bound di errore, e che nella risoluzione dei flussi delle strade non dotate di spira il valore di errore è molto minore.

Poniamoci per semplicità nel caso in cui una spira abbia solo un'altra spira nella sua area di influenza. L'errore sulla spira valutato dal sistema di validazione è un massimo poiché è calcolato tra la spira e la sua vicina; ma nel sistema completo, in cui è di nuovo presente anche la spira rimossa nel calcolo dell'errore, il valore massimo di errore si avrà nel punto a metà delle due spire (si noti che non si parla in questo caso di distanza geografica, ma di una distanza di propagazione dell'informazione sulla rete). In questo punto infatti l'errore proviene da due spire che si trovano a una distanza pari a metà di quella caratteristica dell'errore calcolato, e sarà comunque minore poiché ha due fonti di informazione esatta a distanze minori.

Di fatto molte spire hanno più di una vicina influente, ed ogni vicina ha il suo peso, ma il discorso non cambia sostanzialmente.

Capitolo 4

Analisi Dati

I dati delle spire elaborati nel “Capitolo 2” si presentano nella forma di andamenti giornalieri del flusso corredati da informazioni legate alla loro posizione geografica. Prima di utilizzare questi dati nel modello si è deciso di analizzarli separatamente e poi di cercare legami tra gli stessi.

4.1 Elaborazione sulla media delle spire

Per procedere ad una prima elaborazione dei dati si è scelto di studiare un campione più ristretto di dati, cioè di studiare le spire che avevano valori di flusso più alti; quindi si sono ordinate le spire per flusso medio, così da ottenere un indice della loro popolosità.

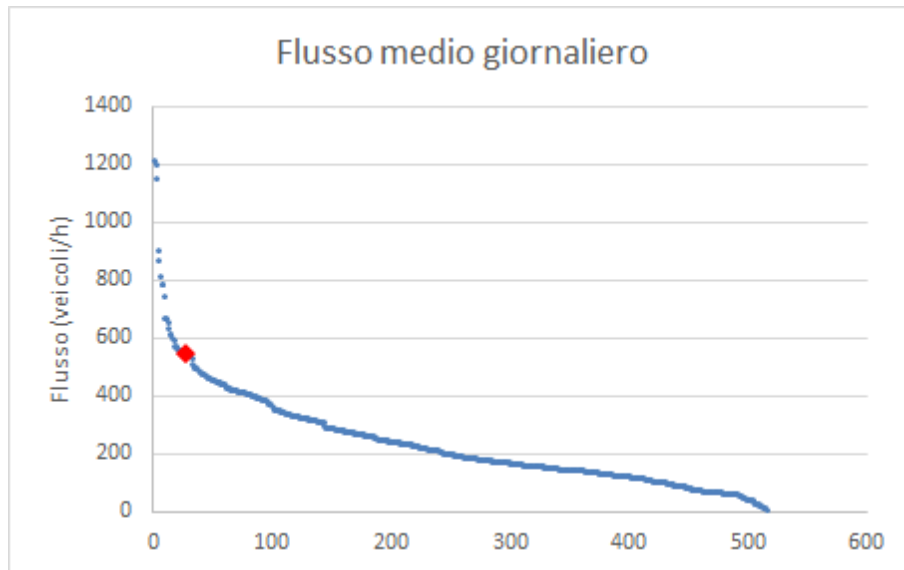


Figura 4.1: è rappresentato il flusso medio giornaliero delle spire, segnando con un rombo rosso la trentesima spira

Inizialmente si è scelto di studiare più a fondo un campione e si sono selezionate le 30 spire più popolate, che sono indicate da quelle a sinistra del rombo rosso in “Figura 4.1”. Questa scelta deriva dal fatto che se un campione è molto popoloso presenta un numero di dati più elevato in tutti i suoi campionamenti orari, ed inoltre si presume di avere una continuità maggiore tra i campionamenti stessi.

Da questa scelta si ottiene il seguente campione di dati.

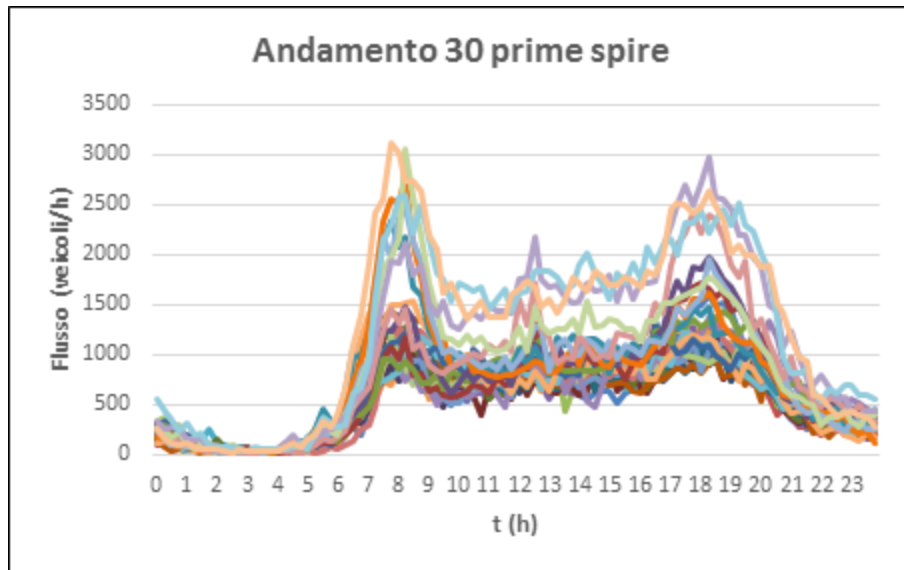


Figura 4.2: sono rappresentate le prime 30 spire più popolate

Per studiare al meglio questo campione si è scelto di calcolarne l'andamento medio ottenendo il seguente grafico:



Figura 4.3: è rappresentata la media degli andamenti delle prime 30 spire più popolate

Per studiare queste spire si è cercato di eliminare la dipendenza dalla loro popolarità normalizzando i vettori flusso. In questo caso la normalizzazione porta i dati ad essere sommati a 96, poiché si è scelto di dividere ogni valore di flusso di ogni spira per il valore di flusso medio.

Dalla normalizzazione si ottiene il seguente grafico:

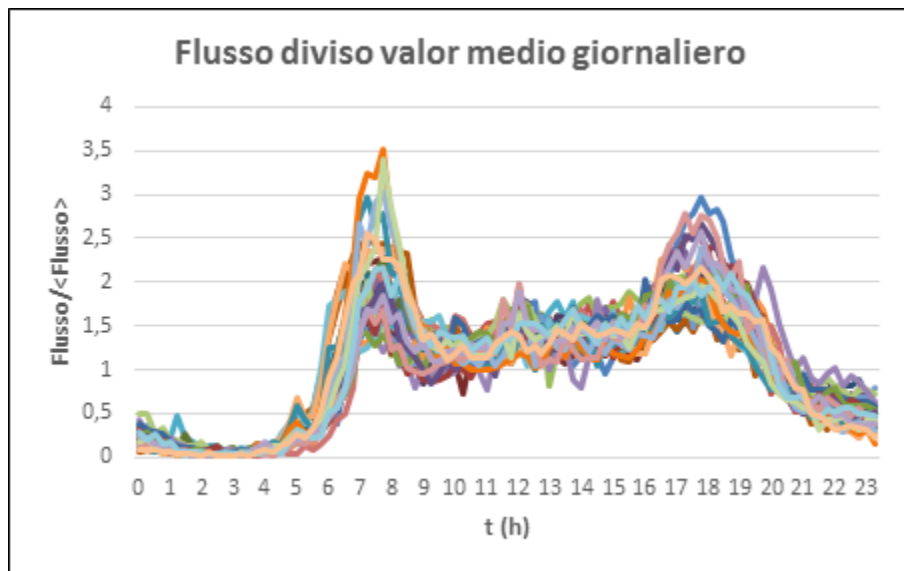


Figura 4.4: sono rappresentate le prime 30 spire più popolose divise per il loro valor medio

Da questi tre grafici si riscontra ciò che si era già osservato nel “Capitolo 3” sugli andamenti giornalieri: il flusso inizia una salita verso le ore 6:00 che porta al picco verso le 8:00, attribuito all’inizio della giornata lavorativa; dalle 9:00 alle 17:00 si mantiene relativamente alto e costante; ha un secondo picco corrispondente al ritorno a casa delle ore 17:00 - 19:00 e si porta lentamente verso lo zero.

Tuttavia tutte queste analisi iniziali non hanno portato risultati particolarmente utili, bisogna infatti far notare che queste operazioni hanno evidenziato l’importanza di considerare il comportamento simile delle spire e quindi le successive analisi hanno mirato in quella direzione.

Per evidenziare le particolarità di ogni spira, infatti, sarebbe desiderabile che questa componente principale non vi fosse. Questa problematica verrà analiz-

zata più avanti nel capitolo, dove si cercherà di studiare gli andamenti senza considerare questa forte componente del sistema, conseguenza della grande isocronia. Si è cercata quindi una funzione in grado di compensare questa necessità, e si è scelto di porre l'attenzione sulla correlazione fra le spire.

Nondimeno, l'osservazione di questo primo campione ha permesso di capire che era necessario studiare un insieme di spire più eterogeneo, dal quale ottenere informazioni più particolari sulla natura delle spire.

4.2 Correlazione

4.2.1 Introduzione

La correlazione statistica fra due variabili dipendenti $x_1(t)$, $x_2(t)$ è una grandezza che misura la tendenza di una variabile a variare al variare dell'altra rispetto al parametro indipendente t . La misura non riguarda la relazione di causa-effetto fra le due variabili, ma semplicemente ne indica la tendenza a variare analogamente. In generale il valore di correlazione assume un valore fra -1 e 1 , dove 1 indica che le variabili sono totalmente correlate, cioè la variazione di una è ricavabile dalla variazione dell'altra; 0 indica che le variabili sono scorrelate e la conoscenza della variazione di una variabile non dà informazioni sulla variazione dell'altra; -1 indica che le funzioni sono totalmente anticorrelate, cioè che hanno variazioni inverse sul parametro.

In figura vengono mostrati diversi esempi di correlazione:

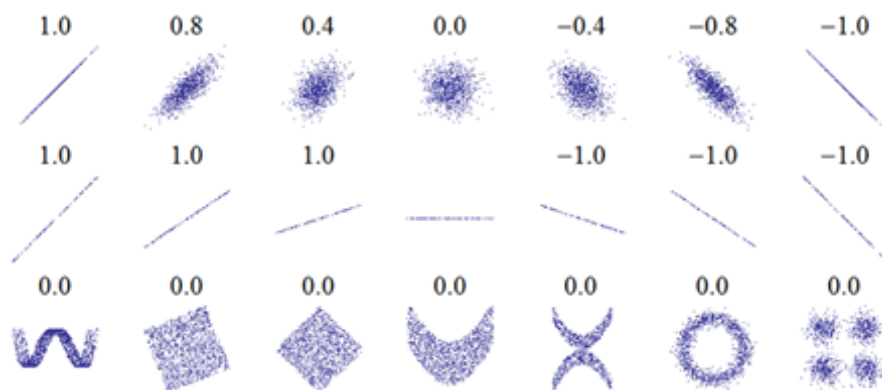


Figura 4.5: sono rappresentati alcuni esempi di punti nello spazio correlati con valore di correlazione

Supponendo di avere due vettori funzione \mathbf{f}_i e \mathbf{f}_j , si possono trattare come due variabili casuali $x_i(t)$ e $x_j(t)$.

Assumendo inoltre che i vettori abbiano 96 componenti, la correlazione viene calcolata nella seguente maniera (per eliminare gli indici è preferibile utilizzare \mathbf{f} e \mathbf{g} , al posto di \mathbf{f}_i e \mathbf{f}_j):

$$corr(\mathbf{f}, \mathbf{g}) = \frac{\sum_{t=0}^{95} [(f(t) - \langle f \rangle)(g(t) - \langle g \rangle)]}{\sqrt{\sum_{t=0}^{95} (f(t) - \langle f \rangle)^2 \sum_{t=0}^{95} (g(t) - \langle g \rangle)^2}} \quad (4.1)$$

Dove $f(t)$ sono i punti della prima funzione $\langle f(t) \rangle$ è il valor medio della prima funzione mentre $g(t)$ e $\langle g(t) \rangle$ sono relativi alla seconda funzione.

Si noti che, definendo $\mathbf{f}^* = \mathbf{f} - \langle f \rangle$, dove la sottrazione è effettuata per ogni componente del vettore, la correlazione si riduce al calcolo del prodotto scalare Euclideo dei vettori star normalizzati:

$$corr(\mathbf{f}, \mathbf{g}) = \langle \widehat{\mathbf{f}^*} | \widehat{\mathbf{g}^*} \rangle \quad (4.2)$$

L'idea alla base dello studio delle correlazioni nel presente lavoro nasce dalla considerazione che sarebbe interessante ottenere funzioni che permettano di osservare legami tra spire vicine e quindi poter migliorare la propagazione rispetto a differenti parametri osservando quanto le variazioni tra i diversi flussi giornalieri siano interdipendenti, evidenziare eventuali ritardi temporali nella correlazione fra due andamenti, trovare fra le spire classi di somiglianza dei flussi, cercare delle relazioni fra distanze geografiche e correlazioni.

Inizialmente, si è associato ad ogni spira un vettore di flusso calcolato come andamento giornaliero medio, sui 20 giorni feriali del mese di maggio 2011, dei flussi di autoveicoli; in questo paragrafo, quando si parlerà di correlazione tra spire, si intenderà la correlazione tra questi vettori così creati.

Il campione consiste di circa 500 funzioni a 96 componenti, avendo un valore ogni quarto d'ora per 24 ore.

4.2.2 Correlazione Triviale

Inizialmente la funzione di correlazione statistica è stata applicata ad ogni coppia di spire per poter osservare come queste variassero tra di loro. Il risultato è stato che le diverse spire in una buona parte delle situazioni presentano un'alta correlazione tra di loro, il ché indica una certa coerenza nella variazione giornaliera. Questo era già stato previsto poiché si erano osservati i diversi flussi

giornalieri di differenti spire e si era osservato come questi variassero con regolarità: il plateau notturno e quello giornaliero ed i picchi mattutini e serali sono comuni a tutte le spire, e per quanto le ampiezze medie non siano uguali per tutti l'andamento rimane comunque simile.

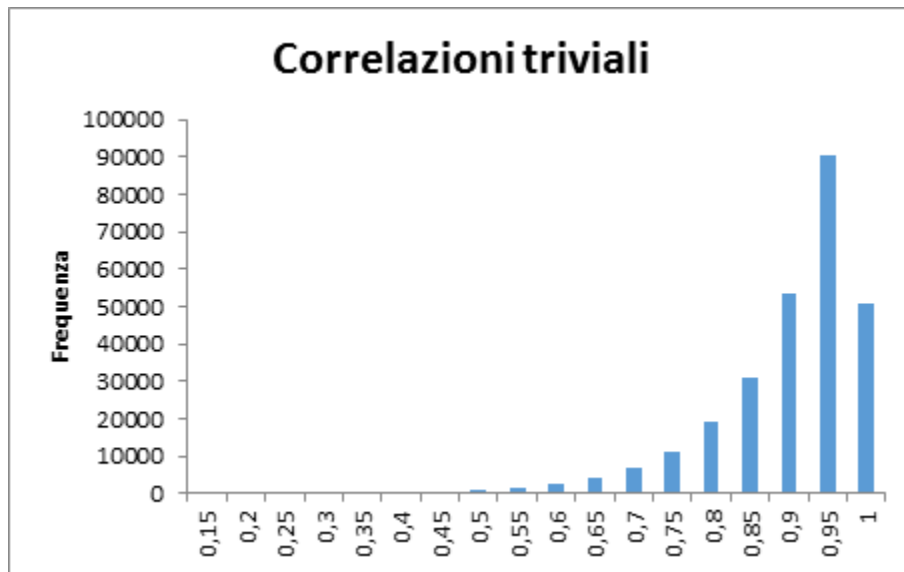


Figura 4.6: sono rappresentati i valori di correlazioni tra tutte le coppie di spire, sotto ad ogni colonna dell'istogramma è presente il valore massimo di ogni bin

Il metodo utilizzato per calcolare la correlazione è quello indicato dalla 4.1 e si può osservare dall'istogramma che si presentano grandi occorrenze per valori di correlazione molto alti, mentre nessuna coppia di spire si mostra anticorrelata.

Un valore così alto di correlazione media tra le spire invece era inatteso. Una spiegazione di questo fenomeno è stata ricercata nella forte isocronia del sistema, cioè la proprietà che il sistema tende a variare in maniera coerente indipendentemente dalla posizione delle spire.

Questo risultato, vista l'alta coerenza del sistema, non porta sufficiente informazione sugli andamenti delle spire per poterle suddividere, quindi si è scelto di cercare di studiare in maniera differente quest'indice di correlazione.

4.2.3 Correlazione modulare triviale

Lo studio successivo è stato quello di cercare un eventuale ritardo temporale

nella trasmissione dell'informazione tra le differenti spire. L'idea era che le correlazioni avrebbero potuto dare risultati più soddisfacenti se calcolate con uno shift modulare sui tempi, in modo che tenessero conto dei ritardi di propagazione e permettessero così di osservare eventuali massimi di correlazione a tempi corrispondenti alle distanze geografiche.

Per questo studio quindi si sono ricalcolate le correlazioni fra ogni spira e ogni traslazione temporale di tutte le altre, secondo la seguente formula

$$\text{corr}(\mathbf{f}, \mathbf{g}, n) = \frac{\sum_{t=0}^{95} [(f(|t+n|) - \langle f \rangle)(g(t) - \langle g \rangle)]}{\sqrt{\sum_{t=0}^{95} (f(|t+n|) - \langle f \rangle)^2 \sum_{t=0}^{95} (g(t) - \langle g \rangle)^2}} \quad (4.3)$$

quindi $\text{corr}(f(|t+n|), g(t))$

Dove il valore n indica la traslazione temporale, ed è stato fatto variare da 0 a 95 ogni volta rendendo la funzione modulare, così da confrontare ogni funzione con tutti gli shift dell'altra.

In questo modo si sono ottenuti circa 30 milioni di valori di correlazione, e per ogni coppia di spire è stato selezionato il valore di spostamento temporale, n , che produceva il massimo di correlazione tra le spire, così sono stati inseriti in un istogramma questi valori ottenendo il seguente risultato:

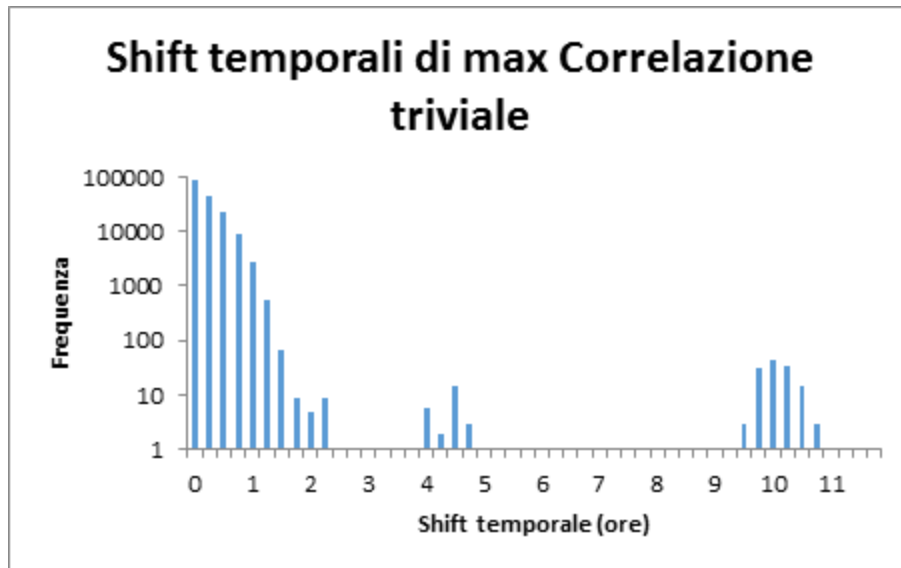


Figura 4.7: sono rappresentati i valori di shift dove viene trovato il massimo di correlazione per ogni coppia di spire, la scala sulle ordinate è logaritmica per permettere di osservare al meglio le svariate occorrenze

Si noti che in questo istogramma si erano ottenuti all'incirca 150 mila dati e che il picco attorno allo 0 arriva a 90 mila occorrenze. Il grafico mostra solamente l'intervallo 0:00 - 12:00 perché per definizione delle coppie di correlazione questo grafico sarebbe simmetrico rispetto a mezzogiorno, perché se spira A ha massimo di correlazione con spira B shiftando a più +7 il tempo, il valore massimo di correlazione di spira B con spira A si otterrà con shift pari a -7 .

È possibile osservare che si ottengono tre zone in cui le occorrenze sono maggiori di 0; la prima zona più evidente è quella centrale attorno all'ora zero, che corrisponde ad un ritardo temporale pari o vicino a 0, il che indica che il sistema, anche se spostato di qualche ora mantiene una certa isocronia; successivamente è presente una zona attorno alle 4:00 di mattina in cui si osserva un picco minore e questo si è spiegato osservando direttamente gli elementi in questione e apprendendo che tutte le correlazioni provengono da una spira, la 1133, che ha un andamento particolare, mostrato successivamente. Invece la terza zona è spiegata dal fatto che, nel calcolare la correlazione migliore shiftando temporalmente, le due spire hanno trovato una maggior correlazione facendo corrispondere il picco pomeridiano con quello mattiniero dell'altra, portando ad osservare uno shift temporale di quasi una metà giornata. Lo shift non è pari a 12 ore, perché i

picchi mattinieri e pomeridiani non sono simmetrici rispetto a mezzogiorno, ma rispetto alle 10:00, come è infatti possibile osservare dal grafico in cui il secondo picco è proprio attorno alle 10:00.

È quindi graficata la spira 1133:



Figura 4.8: è rappresentato il flusso giornaliero mediato sulle giornate feriali del mese di maggio del 2011

Si osserva che questa spira è molto particolare, infatti è priva del picco pomeridiano. Questo ha portato a mostrare una massima correlazione con altre spire a shift di tempi inaspettati.

Di seguito sono mostrati i valori di correlazione media per ogni bin del grafico precedente.

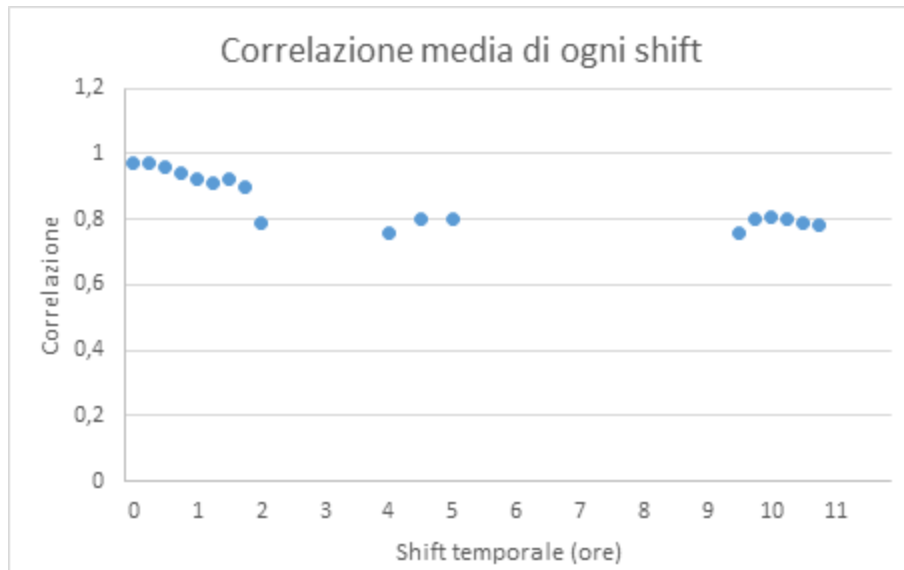


Figura 4.9: sono rappresentati i valori medi di correlazione per ogni bin del grafico delle correlazioni shiftate

In questa seconda figura si osserva, per ogni shift temporale, il valore di correlazione mediato su tutti gli elementi aventi quello shift come massimo. Si può osservare che questo valore di correlazione diminuisce man mano che lo shift massimo si allontana dallo zero.

Il motivo di questo andamento è la grande isocronia del sistema: nei casi in cui il massimo non è a zero, la correlazione rimane comunque più bassa che nei casi, largamente prevalenti, in cui lo shift massimo è quello nullo. Anche questo grafico quindi costituisce una riprova della presenza di una componente giornaliera molto importante e ubiqua.

Successivamente per confrontare i valori dei ritardi temporali tra le spire e le correlazioni shiftate si è scelto di graficare la differenza tra le distanze temporali tra la coppia di spire più influenti calcolate tramite la posizione geografica e il picco massimo dello shift di correlazione:

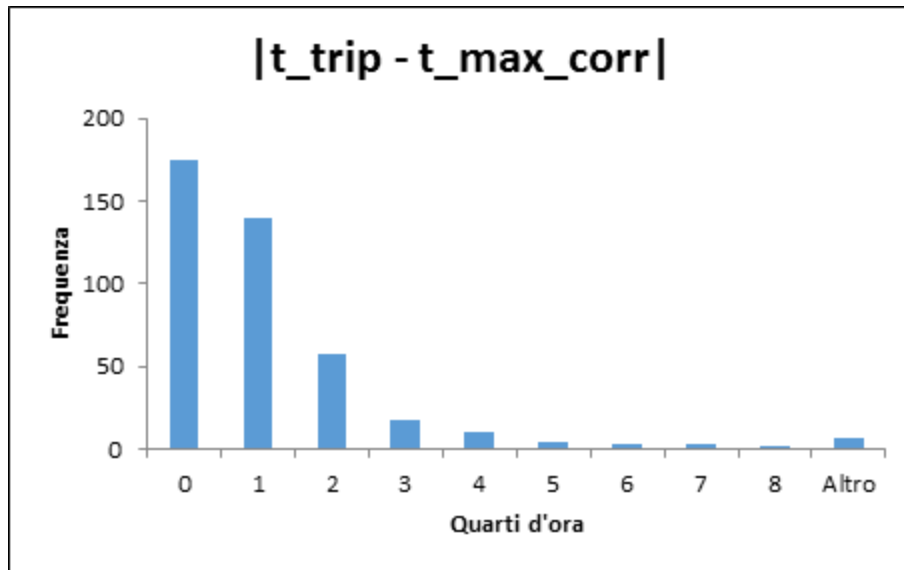


Figura 4.10: sono gli scarti tra il ritardo temporale tra le spire ed il ritardo in cui la correlazione è massima, in quarti d'ora

È possibile osservare che per una gran parte dei dati (circa il 75%) questo ritardo è compreso tra -15 e $+15$ minuti.

4.2.4 Correlazione raffinata

Dopo aver calcolato i valori di correlazione tra le diverse spire ed aver osservato che la componente fondamentale dei flussi è quella isocrona si è cercato di eliminare questa forte componente per rendere più distinguibili gli andamenti.

Prima di procedere all'elaborazione dei dati si è scelto di selezionare una piccola parte del campione dei dati, poiché era interessante osservare l'andamento del campo medio, cioè la componente portante giornaliera solamente di un gruppo di spire con andamenti simili e caratterizzanti del sistema: è stato scelto di escludere le prime 15 spire, perché da come si può osservare dalla "Figura 4.1" mostrato precedentemente hanno valori di flusso molto più alti rispetto alle successive. Quindi sono state selezionate le successive 100 spire, in quanto ancora molto popolose e con valori di flusso medio abbastanza simili, e con questi 100 vettori di flusso è stato calcolato il campo medio, che poi sarà eliminato da tutte le spire per cercare di eliminare la componente isocrona del sistema.

Uno degli obiettivi di questo affinamento nello studio delle correlazione è quello di trovare una classificazione delle spire in termini di correlazione, in cui ogni classe contenga spire tra di loro molto correlate; questo avrebbe infatti permesso di osservare comportamenti di gruppo e visualizzarli anche dal punto di vista geografico.

Le correlazioni si sono ricalcolate associando inizialmente ad ogni spira j il seguente vettore caratteristico φ_j

$$\varphi_j = \mathbf{f}_j^* - \langle \mathbf{f}_j^* | \hat{\mu} \rangle \hat{\mu} \quad \mathbf{f}_j^* = \mathbf{f}_j - \langle f_j \rangle \quad (4.4)$$

Dove:

- f_j è l'andamento della spira j ;
- μ è il campo medio, cioè l'andamento medio delle 100 spire selezionate, e $\hat{\mu}$ la sua normalizzazione;
- $\langle f_j \rangle$ è il valor medio dell'andamento j ;
- di conseguenza \mathbf{f}_j^* è l'andamento j -esimo traslato sull'origine.

Il vettore caratteristico di una spira è quindi la distanza fra l'andamento traslato a zero e la sua proiezione vettoriale lungo il campo medio.

Il nuovo valore di correlazione di una coppia di spire si è ottenuto quindi come prodotto scalare cartesiano fra i due vettori caratteristici normalizzati.

$$corr_{raff}(\mathbf{f}_i, \mathbf{f}_j) = \langle \hat{\varphi}_i | \hat{\varphi}_j \rangle \quad (4.5)$$

Si è inoltre deciso di assegnare ad ogni coppia di spire una misura di distanza di correlazione $DC(\cdot, \cdot)$, calcolata come complemento a 1 della correlazione diviso per due:

$$DC(\mathbf{f}_i, \mathbf{f}_j) = \frac{1 - \langle \hat{\varphi}_i | \hat{\varphi}_j \rangle}{2} \quad (4.6)$$

Poiché i vettori caratteristici vengono normalizzati prima del calcolo, la correlazione può assumere ogni valore nel range $[-1; 1]$; come di consueto, il valore 1 indica correlazione totale, il valore nullo indica correlazione nulla e -1 indica anticorrelazione totale.

Si vede quindi che la definizione di distanza scelta assume valori nell'intervallo $[0; 1]$ e un valore nullo di distanza corrisponde alla correlazione massima, una distanza pari a 1 corrisponde all'anticorrelazione totale, e una distanza di $1/2$ indica l'assenza di correlazione.

Un istogramma dei valori di correlazione per la seconda classe di spire è mostrato di seguito.

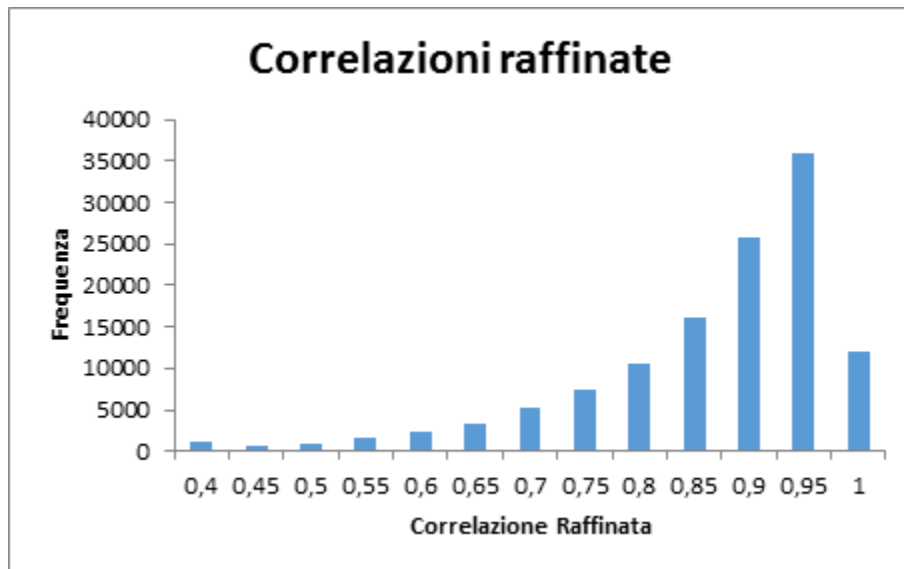


Figura 4.11: è rappresentato l'istogramma delle correlazioni raffinate per ogni coppia di spire, sotto ad ogni colonna dell'istogramma è presente il valore massimo di ogni bin

Nelle analisi e figure successive si è scelto di rappresentare le similarità fra spire in termini della distanza di correlazione piuttosto che di correlazione; questo poiché, come si vedrà, la distanza di correlazione è direttamente collegata all'errore.

Questo istogramma mostra che effettivamente rimuovere la componente principale degli andamenti abbassa le correlazioni del sistema, poiché viene rimossa una componente che accomuna tutte le spire.

Si è inoltre ricalcolato il valore di correlazione shiftata tra tutte le spire selezionate ottenendo il seguente risultato:

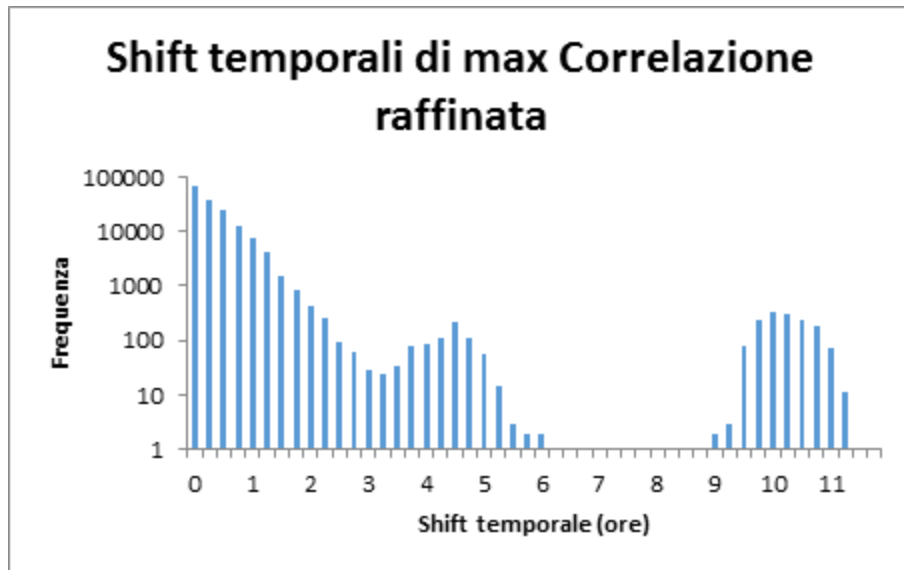


Figura 4.12: sono rappresentati gli shift temporali dei massimi di correlazione per ogni coppia di spire, utilizzando la correlazione raffinata

Si è ottenuto un risultato abbastanza simile a quello ottenuto precedentemente, di nuovo con i tre picchi attorno alle ore 0:00, 4:00 e 10:00. La differenza è che questa tipologia di correlazione cerca di eliminare il campo medio e quindi si osserva che le occorrenze sono meno nette, il che è dovuto al fatto che la componente che rimane ai vettori di flusso giornaliero è stata privata solamente da una parte dell'isocronia e questo porta ad osservare gli stessi picchi, ma in maniera meno netta.

4.2.5 Correlazione tra influenti e massima

Un'ulteriore analisi effettuata è stata quella di confrontare, per ogni spira:

- il valore della sua correlazione con la spira che, nel modello sviluppato, ha più influenza nella propagazione;
- il valore della sua correlazione massima.

La spira più influente è un dato che dipende dalla distanza geografica e dalla natura delle strade che collegano le differenti spire ed è il risultato del processo di propagazione dell'informazione dalle spire.

È importante sottolineare che nella ricerca delle vicine più influenti di ogni spira, sono escluse la spira stessa e la direzione opposta, che portano sempre necessariamente i due pesi maggiori, poiché la loro influenza è triviale e poco interessante.

La formula utilizzata per calcolare la correlazione e la conseguente distanza di correlazione è quella mostrata nel paragrafo precedente.

Si è osservato che, solo nel 4% delle spire, la spira più influente è anche quella che ha il valore massimo di correlazione.

È successivamente mostrato l'istogramma delle differenze di correlazione, calcolate come differenza assoluta fra distanza di correlazione minima (correlazione massima) e distanza di correlazione con la più influente.

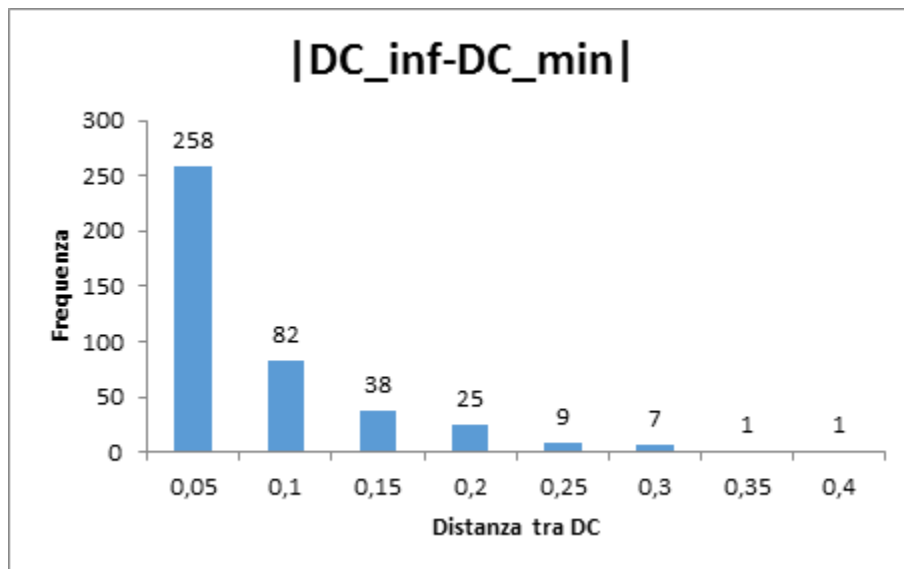


Figura 4.13: sono rappresentate le occorrenze dello scarto tra la distanza di correlazione con la spira più influente con la distanza di correlazione minore

Si osserva un risultato atteso, poiché, anche se il campione di spire per cui la spira di correlazione massima e la più influente coincidono è piccolo, i bin

immediatamente successivi raccolgono un grande numero di elementi, con oltre il 60% di spire entro 0.05 di distanza.

Si può fare un confronto con il grafico seguente, in cui appare la distribuzione dei valori di distanza di correlazione tra ogni spira e la sua più influente.

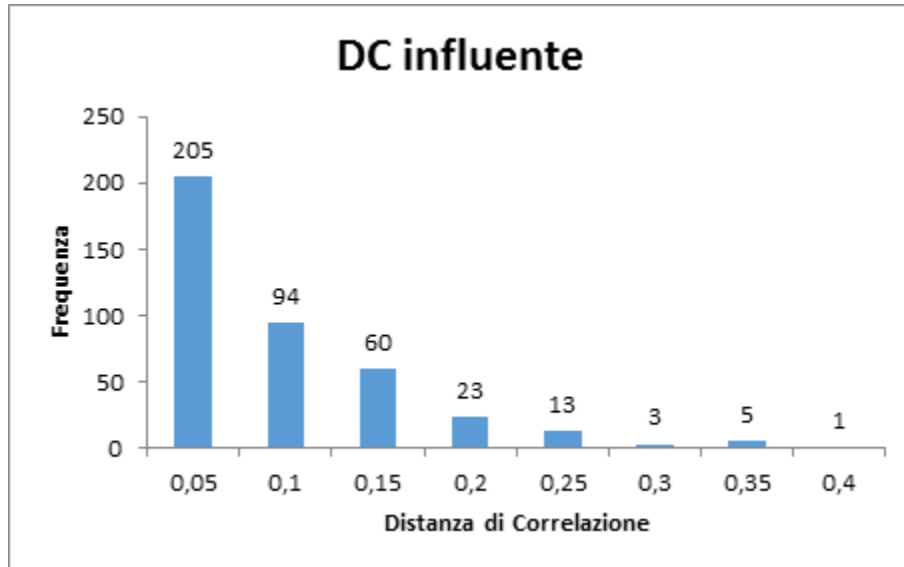


Figura 4.14: sono rappresentate le occorrenze delle distanze di correlazione tra le spire più influenti

Si nota che si può leggere da questo istogramma la stessa informazione di quello appena precedente. Infatti se si osserva una differenza piccola fra correlazione massima e correlazione più influente, è altamente probabile che la spira più influente sia molto correlata.

Siccome correlazione alta significa un valore numerico vicino ad 1, la corrispondente distanza di correlazione varrà circa 0; ciò implica che, assumendo che la distanza di correlazione minima valga circa 0, la differenza $|DC_{influyente} - DC_{min}|$ si riduce al valore di $DC_{influyente}$.

In seguito sono state considerate, per ogni spira, tutte le vicine che la influenzano, calcolando per esse una media pesata della correlazione, ove i pesi sono gli stessi che nel programma misurano l'influenza. Chiamando w_j questi pesi e c_j la correlazione della spira madre con la sua vicina j -esima, si è calcolata la seguente correlazione media:

$$\bar{C}_i = \frac{\sum_j w_{ji} \text{corr}(i, j)}{\sum_j w_{ji}} \quad (4.7)$$

Dove n è il numero di vicine che influenzano la spira i -esima.

Di queste correlazioni si sono poi calcolate le corrispondenti distanze di correlazione, e mettendo questi valori in istogramma si ottiene la seguente figura.

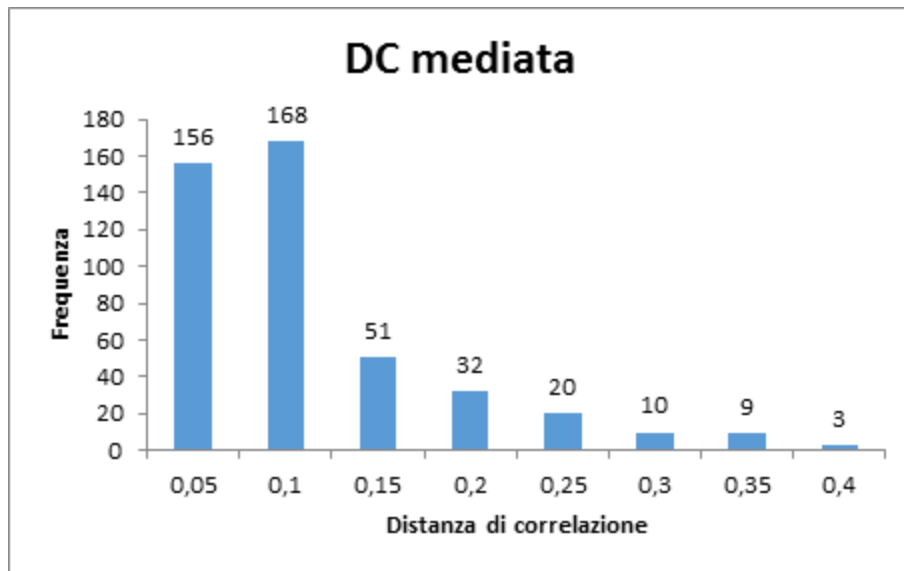


Figura 4.15: sono rappresentate le occorrenze delle distanze di correlazione tra una spira e la media delle sue vicine più influenti secondo i pesi calcolati nella propagazione dell'informazione da parte del programma, escludendo la direzione opposta di ogni spira

Si osserva che, rispetto alla correlazione della prima vicina ove circa il 50% delle spire si manteneva entro lo 0.1, qui quelle entro la soglia sono oltre il 70%. Ciò conferma che utilizzare la media pesata dà risultati migliori rispetto a quella di prima vicina, e vista la natura del sistema di ricostruzione dei flussi diminuisce direttamente l'errore commesso, come si mostrerà in seguito.

Successivamente in “Figura 4.16” si possono confrontare i valori di distanze di correlazione minime con quelle più influenti.

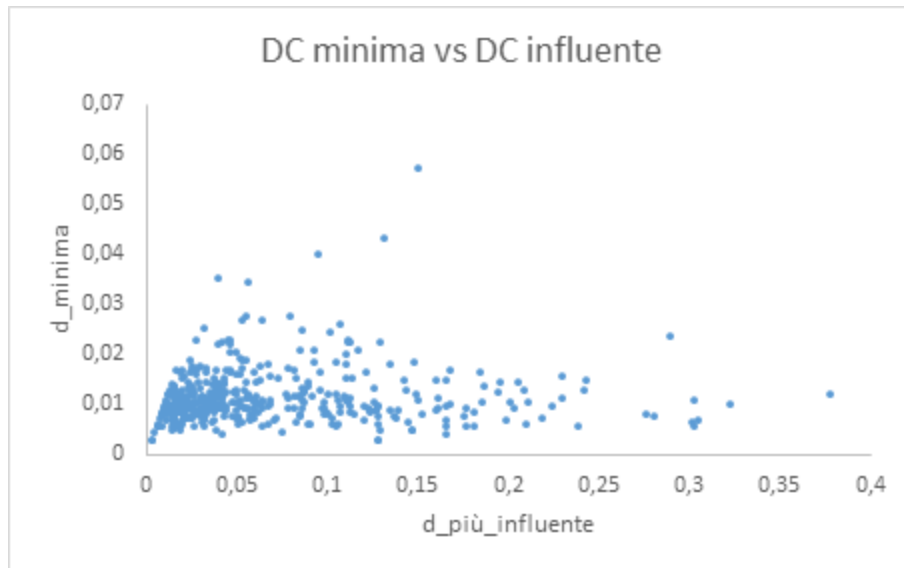


Figura 4.16: sono confrontati i valori di distanza di correlazione tra la spira più influente (in ascissa) e quella più vicina (in ordinata)

4.2.6 Relazioni della correlazione con altre grandezze caratteristiche

Dopo aver osservato che, nella dinamica di una spira, non necessariamente la vicina più influente dal punto di vista della propagazione dei segnali è anche quella con cui la spira originale ha correlazione maggiore, si sono cercate altre relazioni significative fra la correlazione e altre grandezze.

Lo scopo di questa ricerca è di migliorare ulteriormente il modello, poiché avere informazioni a priori sulle relazioni dinamiche e sulle dipendenze fra spire sarebbe di aiuto alla ricostruzione dei flussi

Per prima si è studiata la dipendenza fra l'errore commesso dal modello nel calcolare in flusso su una spira e il valore di distanza di correlazione fra questa spira e la sua vicina più influente.

L'errore è quello calcolato, come esposto nel "Capitolo 3", come media sulla giornata degli errori commessi ogni quarto d'ora nella ricostruzione del flusso sulla spira.

Tra i vari algoritmi di propagazione sviluppati, il più valido in termini di errore finale è quello che utilizza i tempi di transito come pesi e i rapporti fra flussi come fattori di propagazione. In questa analisi è mostrato l'errore compiuto da questa tipologia di algoritmo.

Ordinando le spire in base alla distanza dalla prima vicina, si ottiene la figura seguente:

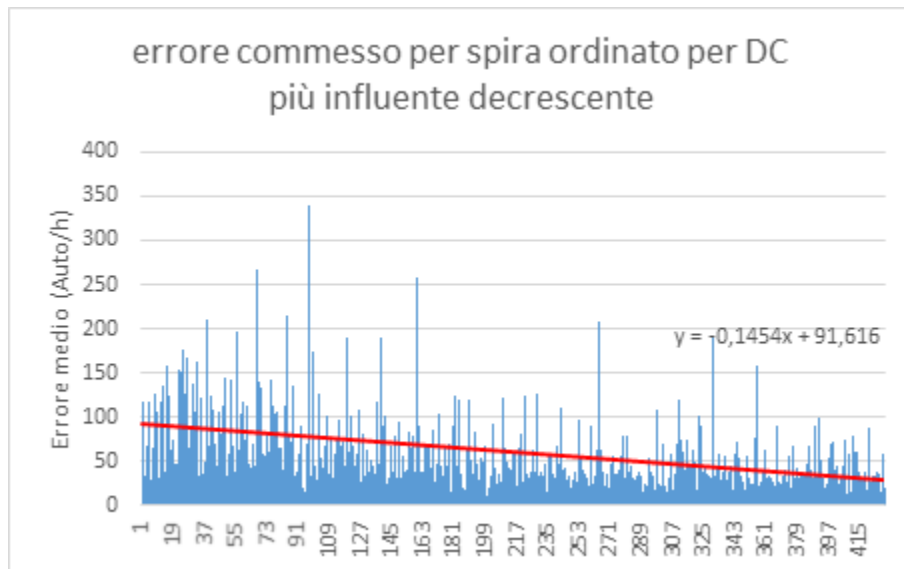


Figura 4.17: sono rappresentati gli errori commessi dal modello ordinando le differenti spire in base al valore di correlazione della più influente

Si osserva ciò che si attendeva, cioè che l'andamento medio dell'errore è affine a quello della distanza dalla prima spira influente, anche se ci si aspettava un andamento più lineare con meno varianza.

Un'analisi più accurata si è ottenuta sostituendo il valore di distanza di correlazione della più influente con il valore mediato delle distanze di correlazioni con tutti i pesi delle differenti spire così da simulare meglio il valore di errore nella validazione del programma.

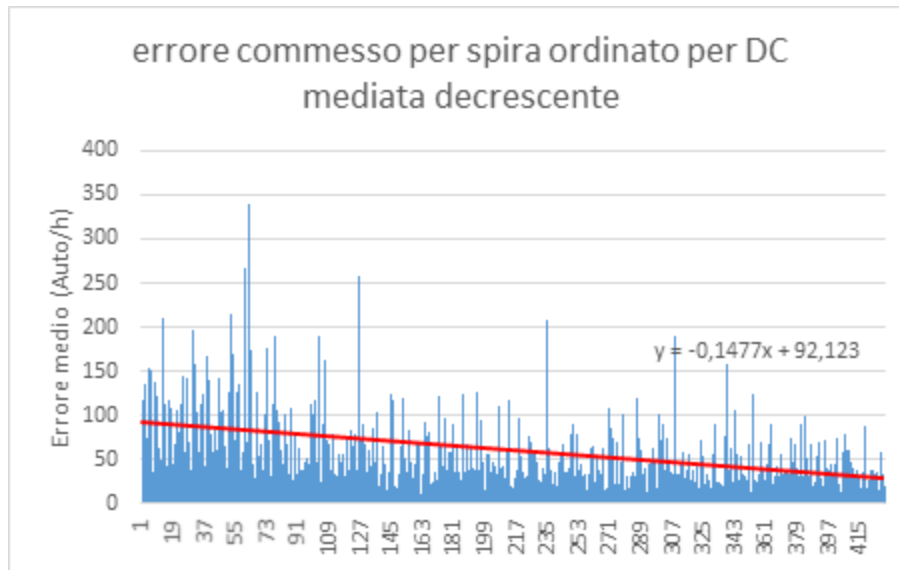


Figura 4.18: sono rappresentati gli errori commessi dal modello ordinando le differenti spire in base al valore di correlazione della media delle influenti

Si ottiene la “Figura 4.18”, in cui di nuovo i dati sono in ordine discendente rispetto al valore di distanza di correlazione mostrando l’errore commesso spira per spira. È visibile l’andamento in discesa: all’aumentare della correlazione media con le vicine, l’errore diminuisce, anche se nuovamente l’andamento della funzione si mostra abbastanza discontinuo. Confrontando i risultati dei due andamenti si osserva che la discesa del grafico con i valori di correlazioni mediate è più rapida rispetto all’altro, quindi è confermato che l’utilizzo di tutto l’insieme delle spire che influenzano quella presa in analisi permette di ottenere un risultato migliore che considerando solamente la prima vicina.

Si è posta l’attenzione su alcuni spike di errore presenti nella parte destra del grafico, dove la distanza di correlazione ha mediamente valore basso, e si è osservato che una buona parte di questi spike è dovuta al fatto che, nel calcolare le ricostruzioni di flusso di queste spire, si compie un’operazione di estrapolazione da spire con flussi più bassi. Quest’operazione è più difficile dell’interpolazione di una funzione tra sue simili, poiché si cerca di trovare il valore di una spira dominante nella zona con funzioni che sono sue subordinate.

Quindi questo errore non è troppo serio a livello di errore totale del sistema, poiché è vero che se quella spira non ci fosse non si saprebbe come vari l’andamento

in quella zona, ma poiché le spire sono state inserite nella regione seguendo una logica legata all'importanza delle strade, non si è troppo preoccupati da questo problema.

4.2.7 Gli errori di ricostruzione

In un'ulteriore analisi ci si è domandati come mai ci siano, fra gli errori, degli spike anomali. Si sono prese quindi in considerazione due particolari spire, una per la quale l'errore è basso e la correlazione media è alta, e l'altra, "anomala", in cui si riscontra un errore alto pur con correlazione media alta.

Per le due spire si sono osservati gli andamenti del giorno 4 maggio 2011, di cui si compie la simulazione; assieme a questi andamenti, si mostrano su grafico tutti gli addendi della sommatoria che nel modello contribuiscono alla ricostruzione dell'andamento, nonché la ricostruzione finale stessa.

Spira	175	152
Errore di ricostruzione	38 Auto/h	173 Auto/h
Distanza correlazione mediata	0.015	0.045
N spire influenti	6	8

Tabella 4.1: sono rappresentate due spire, una anomala cioè che commette un grande errore rispetto al suo valore di distanza di correlazione mediata basso ed una invece con basso errore e bassa distanza di correlazione

4.2.7.1 Spira 175, errore basso e distanza di correlazione bassa

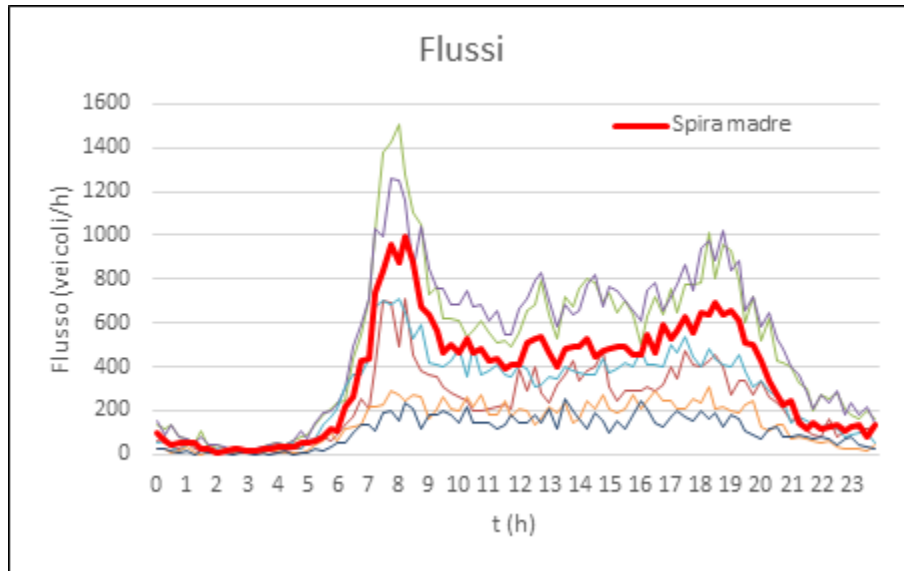


Figura 4.19: è rappresentata la spira 175 in rosso e le sue influenti

In questo grafico appaiono gli andamenti della spira madre, evidenziato in rosso, e di tutte la spire che compongono la sua influenza. Si vede bene che la spira madre ha un flusso che non è mai maggiore della sua vicina più popolosa.

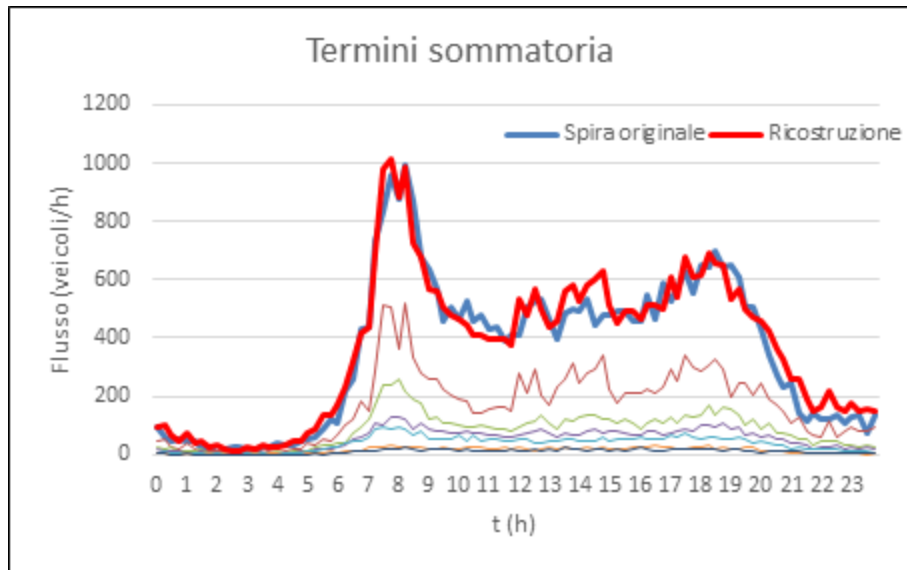


Figura 4.20: questo grafico invece serve a spiegare il metodo di ricostruzione delle spire utilizzato dal metodo di validazione infatti ogni spira è moltiplicata per il vale di peso che ha nella ricostruzione, ed in rosso è rappresentata l'effettiva ricostruzione della spira 175

In questa seconda figura gli andamenti delle spire influenti sono stati moltiplicati per i pesi che hanno nell'algoritmo di propagazione. In sostanza si vedono i termini della sommatoria che ricostruisce il flusso. La linea azzurra spesso rappresenta l'andamento misurato dalla spira, e la linea rossa la sua ricostruzione.

Si osserva come l'errore, che visualmente è rappresentato, a meno di una costante moltiplicativa, dall'area compresa fra le due curve, è piccolo, poiché la ricostruzione è relativamente fedele.

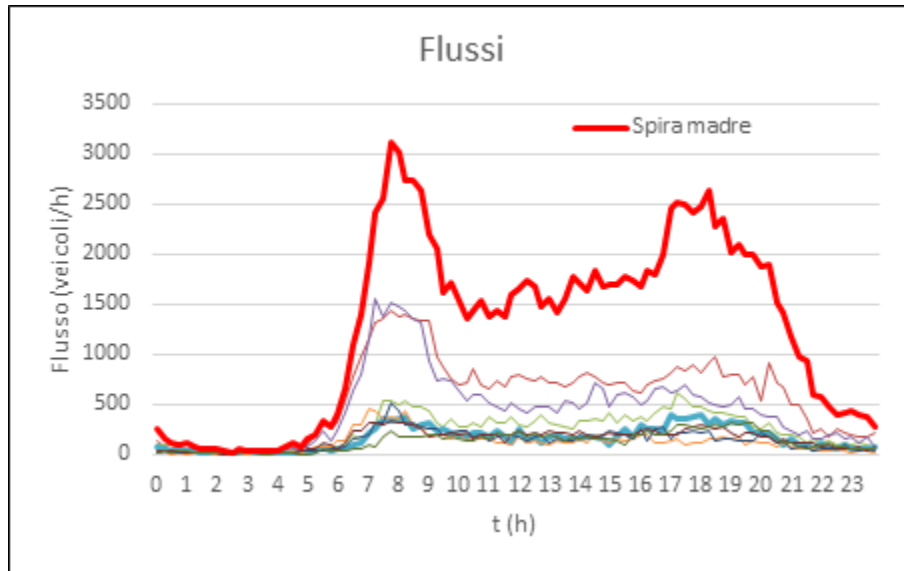
4.2.7.2 Spira 152, errore alto e distanza di correlazione bassa

Figura 4.21: è rappresentata la spira 152 in rosso e le sue influenti

Questo grafico è esattamente analogo alla “Figura 4.19”, ma è relativo alla spira che definiamo anomala a causa dell’alto errore. Di nuovo in rosso è evidenziato l’andamento della spira madre, e in questo caso è evidente che essa ha flussi molto maggiori delle sue vicine influenti. In effetti, il suo flusso è costantemente circa il doppio della sua vicina più popolosa.

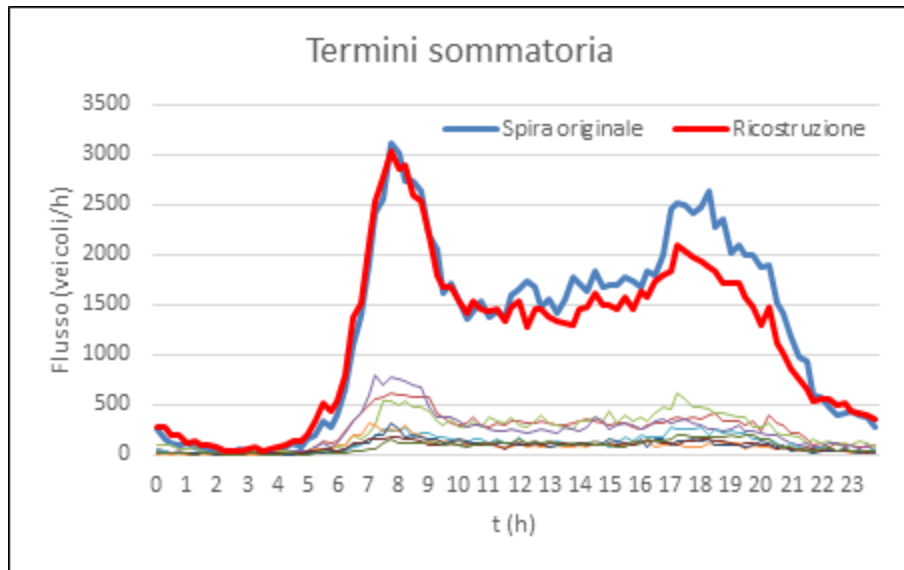


Figura 4.22: analogamente al caso della 175, viene mostrata la ricostruzione della spira 152, in rosso la spira ricostruita, in blu la spira stessa e le spire influenti moltiplicate per i loro pesi

Questo grafico è invece l'analogo a quello mostrato per la spira 175 e mostra gli andamenti ricalibrati dai pesi, cioè i termini della ricostruzione. Qui l'errore commesso nel simulare la spira è alto.

Apparentemente l'area tra le due curve non è eccessivamente grande, sebbene dalle ore 12:00 si osservi uno scarto definito. Bisogna però aver presente che l'errore è calcolato come differenza assoluta e quindi a flussi così alti tale differenza è notevole.

4.2.7.3 Conclusioni

Da questa breve analisi si può ipotizzare che gli errori alti si verificano, come già si era accennato, quando la ricostruzione del flusso deve passare per un processo matematico di estrapolazione. Questo si verifica appunto quando la spira madre è molto più popolosa delle sue vicine.

A tal proposito si mostrano due ingrandimenti della "Figura 4.18" dell'errore di ricostruzione.

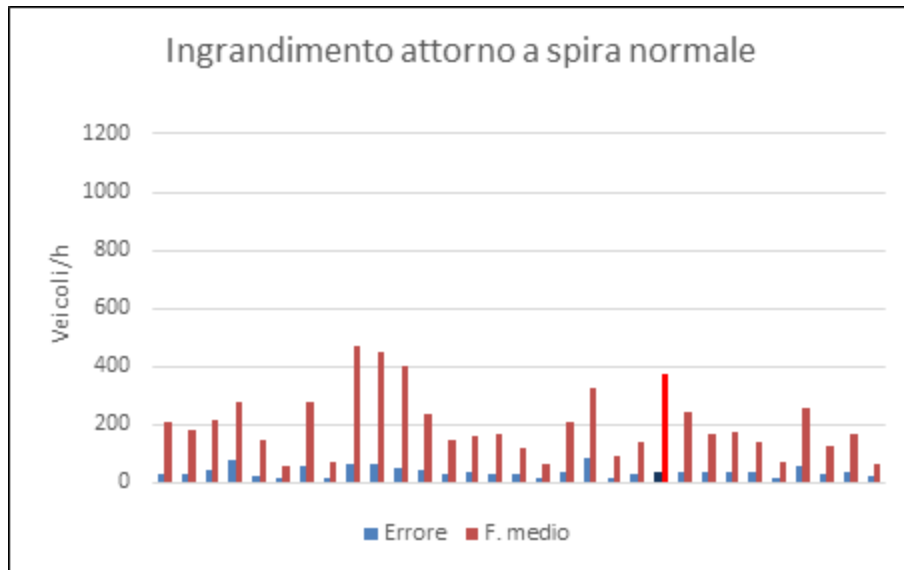


Figura 4.23: è rappresentata una frazione del grafico che mostra l'errore al variare della distanza di correlazione decrescente, in tonalità accese è mostrata la spira 175, sono presenti i valori di errore e di flusso medio orario

Sono evidenziati in colore più brillante l'errore e il flusso medio della spira che si è presa come esempio di spira normale. Il flusso medio è sotto alle 400 auto/h.

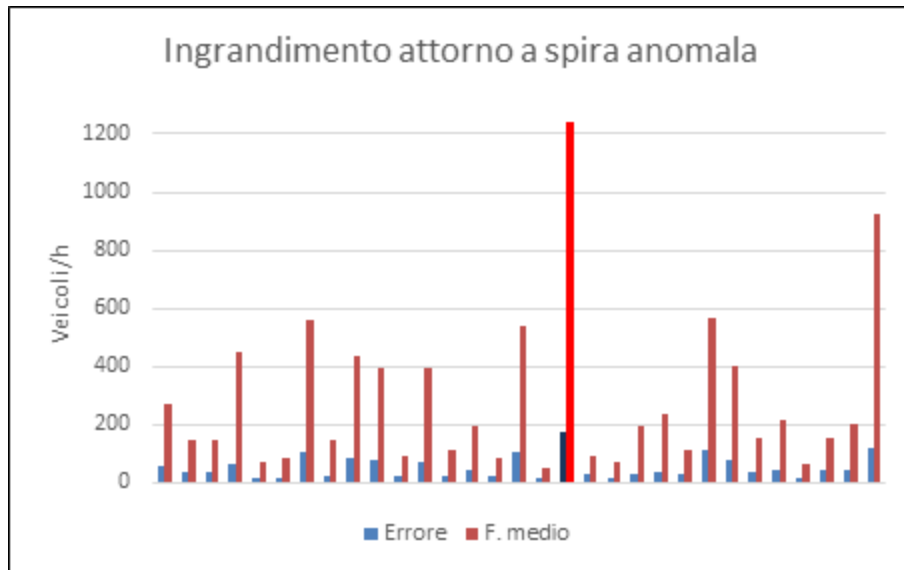


Figura 4.24: analogamente al precedente è rappresentata una frazione del grafico che mostra l'errore al variare della distanza di correlazione decrescente, in tonalità accese è mostrata la spira 152, sono presenti i valori di errore e di flusso medio orario

Con la stessa scala del grafico precedente, è qui mostrato l'ingrandimento attorno alla spira anomala. Si vede che il suo flusso medio è estremamente maggiore della spira normale, e molto alto rispetto alla media delle spire in questa zona di distanza di correlazione.

4.2.8 Divisione delle spire in cluster

Osservato con più precisione le spire e confrontandole con le altre ci si è chiesti se fosse possibile trovare un modo per raggruppare le spire in base a comportamenti comuni.

Per trovare una simile classificazione, si è scelto di clusterizzare il set di spire in base alla loro distanza di correlazione, in modo da ottenere gruppi contenenti ognuno spire con andamenti di flusso fra loro ben correlati.

Nel determinare i criteri di clustering si sono adottati i due metodi di calcolo della correlazione già esposti:

- il primo è quello che utilizza la correlazione triviale, cioè: $corr_{triv}(\mathbf{f}_i, \mathbf{f}_j) = \langle \widehat{\mathbf{f}}_i^* | \widehat{\mathbf{f}}_j^* \rangle$ da cui la distanza di correlazione $DC_{triv}(\mathbf{f}_i, \mathbf{f}_j) = \frac{1 - \langle \widehat{\mathbf{f}}_i^* | \widehat{\mathbf{f}}_j^* \rangle}{2}$
- il secondo è quello utilizzato per la correlazione raffinata dove per trovare il vettore caratteristico al flusso medio mensile è stato eliminato il campo medio e la media del flusso: $corr_{raff}(\mathbf{f}_i, \mathbf{f}_j) = \langle \widehat{\varphi}_i | \widehat{\varphi}_j \rangle$ dove la distanza di correlazione è, analogamente, $DC_{raff}(\mathbf{f}_i, \mathbf{f}_j) = \frac{1 - \langle \widehat{\varphi}_i | \widehat{\varphi}_j \rangle}{2}$

Ottenuti questi valori di distanze di correlazione, i primi due passi del procedimento di clustering sono i seguenti:

- si è scelta, come spira di partenza, la spira più correlata (trivialmente) al campo medio;
- si sono ordinate le spire successive scegliendo di volta in volta la spira più vicina a quella precedente, escludendo ovviamente le spire già enumerate, fino ad esaurirle tutte.

Si sono così ottenuti i seguenti grafici in cui sono rappresentate le spire con i loro valori di distanza dalla spira immediatamente precedente nel ranking.

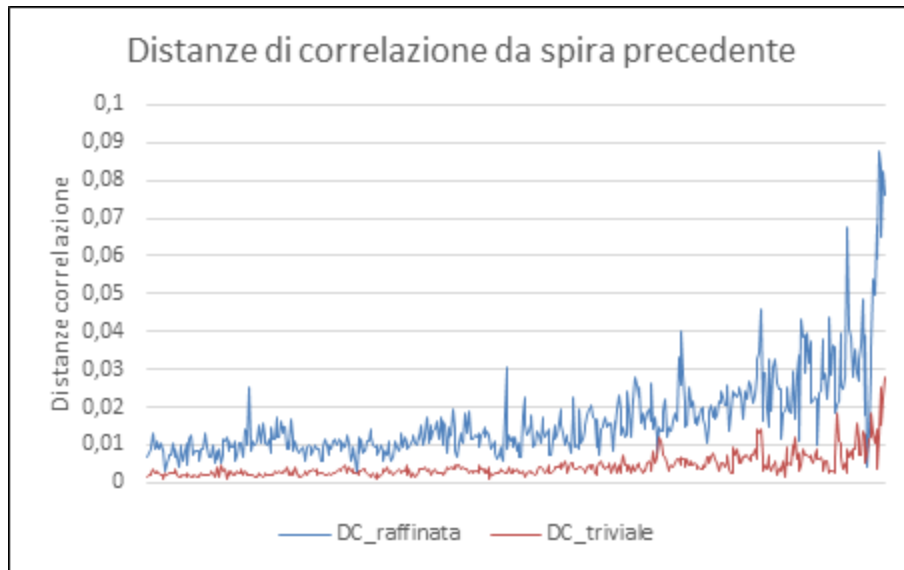


Figura 4.25: sono mostrate le distanze di correlazione ordinate come espresso precedentemente in modo da ottenere i valori per la clusterizzazione. In rosso sono presenti i valori relativi alla distanza di correlazione triviale mentre in blu quelle relativi a quella raffinata

Si nota, risultato già osservato, che utilizzando le correlazioni triviali le spire risultano molto più correlate, mentre utilizzando quelle raffinate i valori di distanza di correlazione sono mediamente più alti.

Per completare il processo di clustering si è scelta una soglia di distanza per separare le classi, in modo tale che ogni cluster raccolga le spire con valore di distanza dalla precedente minore della soglia. Si sono adottati inoltre ulteriori criteri per la divisione:

- si è scelta una soglia iniziale in modo che il primo cluster contenesse all'incirca lo stesso numero di elementi con entrambe le definizioni di distanza utilizzate;
- si è scelto di non creare cluster con pochi elementi e quindi gruppi di poche spire vengono aggiunte al cluster precedente o successivo in base al valore della distanza di correlazione di soglia;
- si è scelto di ottenere una divisione in 6 clusters.

Alla fine del processo, la soglia di distanza per il metodo con correlazione triviale è 0.01, mentre per la correlazione raffinata è 0.035. Questi criteri hanno portato ad ottenere, in entrambi i casi, un primo grande cluster contenente circa 350 spire ed un secondo con circa 60 spire.

Il risultato del clustering effettuato in base alla correlazione triviale è il seguente:

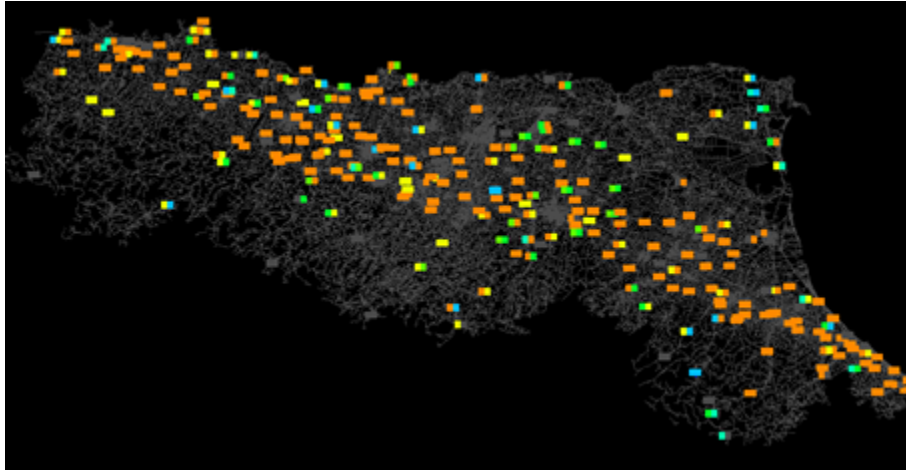


Figura 4.26: è mostrato il risultato del clustering utilizzando i valori di distanze di correlazioni triviali

Dalla correlazione raffinata si ottiene invece la seguente classificazione:



Figura 4.27: è mostrato il risultato del clustering utilizzando i valori di distanze di correlazioni raffinate

4.3 Parametri di influenza e bontà 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.

4.3.1 La bontà

La bontà è definita come

$$B_i = \sum_j \frac{w_{ij} \text{corr}_{raff}(i, j)}{w_{ij}} \quad (4.8)$$

Dove $\text{corr}_{raff}(i, j)$ è la correlazione raffinata calcolata come precedentemente descritto nel capitolo precedente dalla 4.5, mentre w_{ij} indica il peso con cui la spira i influenza la spira j .

Questa sommatoria infatti non viene calcolata sulle spire che compongono l'influenza della spira i -esima, bensì, al contrario, sulle spire che sono influenzate da essa. Di fatto, quindi, la bontà è un parametro che indica quanto una spira venga "chiamata" nel programma e quanto bene questa venga sfruttata, calcolato come media pesata delle correlazioni con le spire influenzate.

Infatti, ricordando che la correlazione vale da 0 a 1, e assumendo il caso estremo che la spira in questione abbia correlazione massima 1 con tutte le spire che la chiamano, la bontà stessa assumerebbe il valore massimo di 1. All'estremo opposto, se la spira fosse completamente scorrelata dalle sue chiamanti, avrebbe bontà nulla.

La media pesata significa inoltre che spire chiamanti poco correlate ma con basso peso non influiscono troppo sulla bontà.

Ecco la distribuzione dei parametri di bontà delle spire mostrati nella seguente figura:

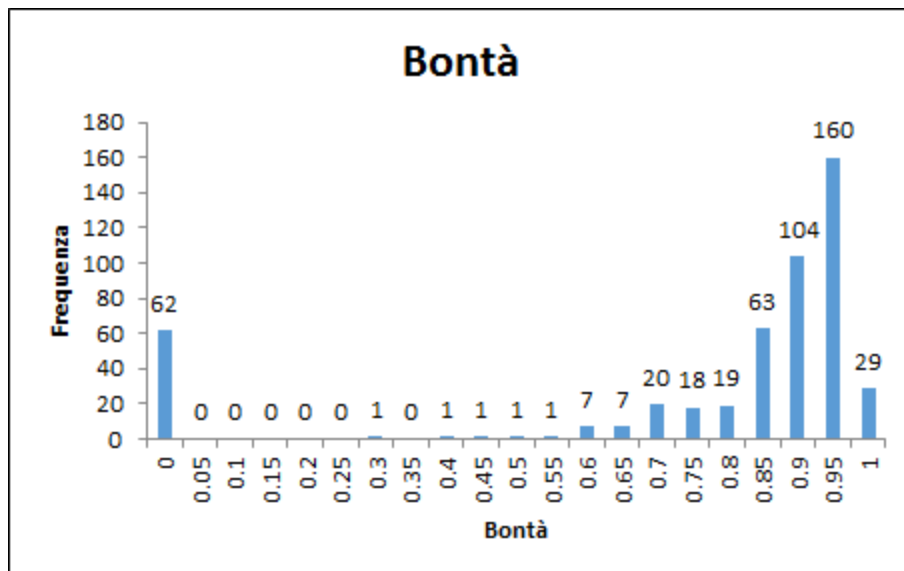


Figura 4.28: sono mostrate nell'istogramma le occorrenze dei valori di bontà per le spire considerate. Si precisa che il bin 0 ha misura inferiore rispetto agli altri valori, ma il suo significato è matematico e non fisico.

Si può osservare dall'istogramma le occorrenze del parametro di bontà per le differenti spire e si consideri che nel bin che ha come valore massimo 0 ha un significato teorico poiché rappresenta le spire che nella validazione non sono utilizzate da altre nel calcolo della ricostruzione dei flussi. Inoltre si consideri che questo parametro per spire con valore pari a 0 determina semplicemente che la spira in questione non viene utilizzata da altre nel programma e non è un indicatore di comportamenti anomali.

Si osserva dall'istogramma che questo parametro assume spesso valori alti, il ch  indica che mediamente quando una spira ne chiama un'altra la correlazione tra queste   alta.

Inoltre il denominatore della frazione, vale a dire la somma dei pesi con cui   chiamata la spira,   a sua volta un indice a sua volta importante, poich  rappresenta una misura dell'utilizzo della spira all'interno del sistema.

4.3.2 L'influenza spaziale

L'influenza spaziale   sostanzialmente la misura del "chilometraggio" totale dell'informazione proveniente dalla spira. Formalmente   definita come la somma delle lunghezze delle strade su cui l'informazione diffonde:

$$I_i^{(s)} = \sum_j w_{ij} l_j \quad (4.9)$$

Dove w_{ij}   il peso con cui l'influenza della spira i arriva in quella determinata strada, mentre il secondo termine l_j   semplicemente la lunghezza di quella strada.

La scelta di inserire la lunghezza in questo modo   dovuta al fatto gi  accennato che la cartografia spesso divide una strada in differenti tratti, ed in questo modo ogni segmento pu  essere considerato separatamente senza andare ad aumentare erroneamente l'indice, poich  questo parametro   la somma della lunghezza di quel pezzo di poligonale con il peso nello stesso punto.

4.3.3 Influenza di flusso

  un parametro molto simile a quello calcolato precedentemente ed infatti   l'indice della quantit  di flusso che quella determinata spira governa nel suo intorno.

  definita in maniera analoga all'influenza spaziale:

$$I_i^{(f)} = \sum_j w_{ij} \frac{\overline{\varphi_gps_j}}{p_gps_j} l_j \quad (4.10)$$

Dove l'unico termine che differisce dalla definizione precedente   il rapporto fra $\overline{\varphi_gps_j}$, che   il flusso medio giornaliero di quella determinata strada calcolato da dati GPS, e p_gps_j , che   il termine di penetrazione della strada; il valore di questo rapporto   quindi il reale flusso medio della strada.

Capitolo 5

Mattina e Pomeriggio, le differenze

5.1 Introduzione

Dopo aver analizzato i dati trattandoli come un unico insieme, cioè considerando che proprietà quali penetrazione e percentuali di svolta non variassero sensibilmente durante la giornata, ci si è proposti di mettere in discussione quest'assunzione; quindi si è cercato di osservare se queste grandezze rimanessero effettivamente costanti, o in caso contrario quanto variassero durante la giornata.

L'idea alle spalle di questa ricerca è quella di osservare se, suddividendo il dataset totale in due separati, si possano migliorare i parametri del modello per migliorare le previsioni del sistema. Questa ipotesi deriva dal fatto che tra mattina e pomeriggio e soprattutto durante le ore di punta ci siano comportamenti molto differenti, infatti la tendenza alla mattina degli individui è quella di recarsi al lavoro, mentre durante le fasce serali vengono svolte una gran parte di attività e ci si aspetta che alcune strade siano trafficate molto più di altre in questi due momenti della giornata, quindi è plausibile pensare che alcune zone diventino dei centri di attrazione in alcuni orari rispetto ad altre.

Inoltre, è già stato mostrato che si sono osservate delle differenze di transito nei profili giornalieri anche nelle due corsie di alcune spire, quindi quegli indizi

hanno spinto a portare avanti un'analisi di questo tipo.

Come già detto precedentemente la giornata è stata divisa in due fasce, attorno alle ore di punta del traffico, che si trovano tra le 7:00 - 9:00 per quanto riguarda la mattinata ed attorno alle 17:00 - 19:00 per il pomeriggio.

In questo capitolo sono illustrate le analisi, dal punto di vista della divisione in fasce, di tre parametri:

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

5.2 Il confronto tra le penetrazioni

Le penetrazioni sono già state calcolate in precedenza e per valutare la variazione di questo parametro si è scelto di utilizzare la penetrazione che definiamo esatta, cioè il valore calcolato tra veicoli che utilizzano la strumentazione GPS e i veicoli transitanti sulle spire.

Per ottenere i valori di mattina e pomeriggio sono stati rielaborati i dati che contenevano i valori delle percorrenze dei GPS dei 20 giorni feriali del mese di maggio del 2011 selezionando solo le opportune fasce orarie e confrontando coerentemente questi valori con quelli delle percorrenze sulle spire.

Così si sono ottenuti due nuovi set di dati, i quali sono stati confrontati tra di loro. Ogni set di dati è composto di 523 valori, che rappresentano le corsie delle strade dotate di spire con informazioni utili.

Si è studiato quindi il seguente rapporto:

$$\rho_i = \frac{|p_gps_i^{(matt)} - p_gps_i^{(pom)}|}{p_gps_i^{(giornata)}} \quad (5.1)$$

Calcolando una distribuzione di questo valore si ottiene la seguente figura:

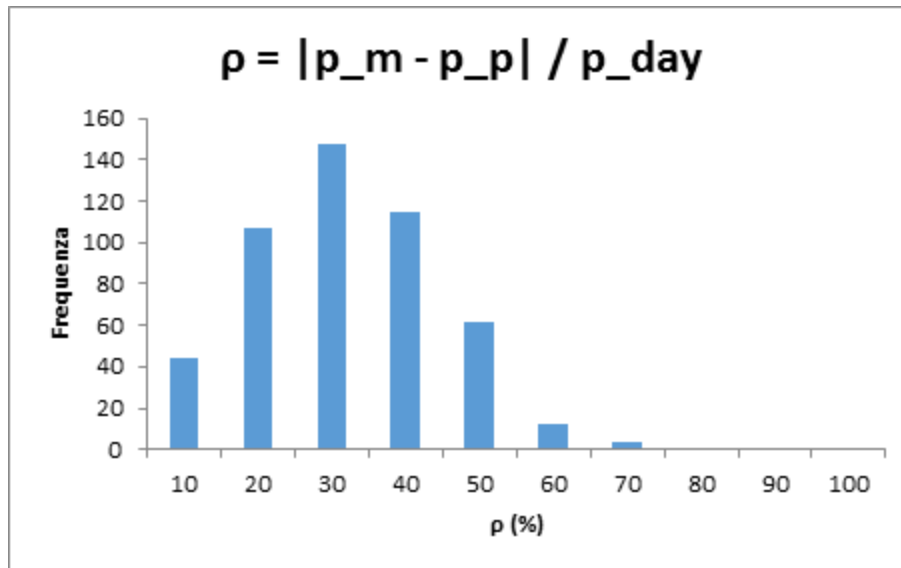


Figura 5.1: sono mostrate le occorrenze del rapporto considerando che il valore sotto al bin dell'istogramma ne indica il limite superiore

Questa figura mostra che la distribuzione esibisce un picco attorno al valore 30% di variazione percentuale ρ . Questo valore è ritenuto più alto di quanto ci si aspettasse, e suggerisce che un'analisi di questo tipo sia importante.

5.2.1 Divisione in tre fasce

A partire da questi risultati si è suddivisa la variazione totale di penetrazione in tre fasce per ricercare come questa variazione si rifletta sulla geografia dell'intera regione e per evidenziare se ci siano regioni con comportamenti coerenti, cioè aree in cui la variazione fra mattina e pomeriggio è simile, aree con variazione alta ed aree con variazione bassa.

Il binning è stato effettuato in base alle occorrenze dei differenti valori, in modo da ottenere popolazioni di ugual numero in ogni bin. I bin risultanti sono quelli per cui il rapporto 5.1:

- è inferiore a 0.212;
- è compreso tra 0.212 e 0.322;
- supera il valore 0.322.

Questo è il risultato sulla cartina.



Figura 5.2: è mostrato il risultato su cartina della divisione in tre fasce

Ogni spira è rappresentata da un rettangolo composto di due colori per indicare le due opposte corsie; i rettangoli in grigio rappresentano le spire scartate per patologie proprie e/o mancanza di dati.

Si può osservare che le spire che presentano basse variazioni, identificate in rosso, sono abbastanza sparse e non direttamente vicine a centri nevralgici, evidenziando che la percorrenza su quelle strade non varia nelle differenti ore della giornata.

Di fatto non si osserva una rilevante coerenza spaziale nei valori di variazione e non si è stati in grado di osservare una regola che governi queste variazioni.

Il passo seguente è stato quello di studiare la distribuzione delle coppie di colori per i rettangoli delle spire, cioè un'osservazione sulle correlazioni nelle variazioni dei due versi di percorrenza delle spire.

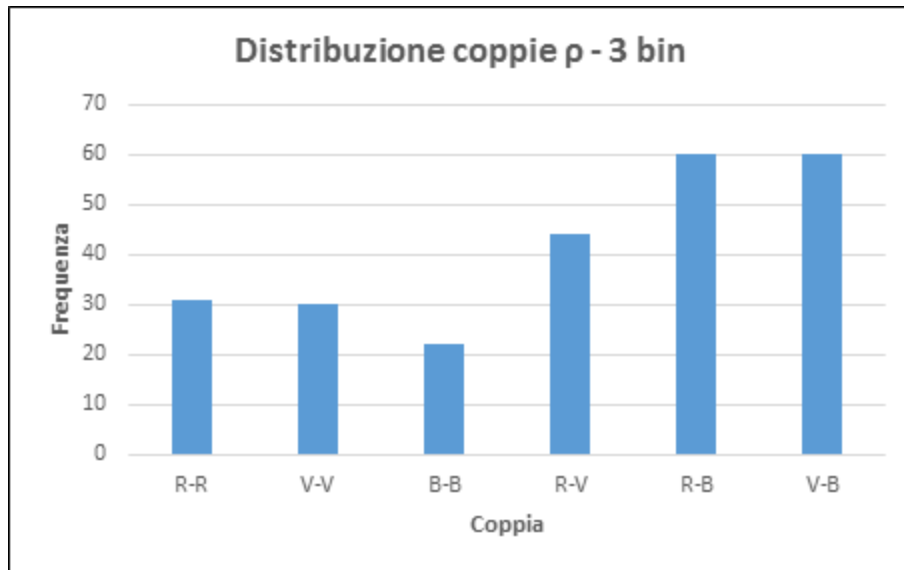


Figura 5.3: è mostrato l'istogramma delle occorrenze delle tre fasce, ogni bin rappresenta una determinata coppia di range di valori.

Si osserva che le combinazioni R-R e V-V hanno circa la stessa frequenza, mentre quella B-B (in cui le due corsie della strada con spira hanno entrambe errore maggiore di 0.322) è meno frequente. Fra le coppie disomogenee, R-V è quella meno frequente, fatto ovvio poiché ogni colore ha un ugual numero di elementi, e R-B e V-B hanno uguale frequenza.

In ogni caso da questo istogramma e dalla rappresentazione su mappa non è possibile osservare direttamente una natta suddivisione in zone; quindi si è cercato di capire se la divisione in tre fasce fosse una frammentazione troppo grande, e si è diviso il set iniziale in soli due range.

5.2.2 Divisione in due fasce

Anche in questa ulteriore suddivisione si è scelto di lasciare uguale la cardinalità dei due insiemi, ottenendo una divisione del campione in una fascia minore ed una maggiore del valore di soglia $\rho < 0.272$ dalla 5.1 :

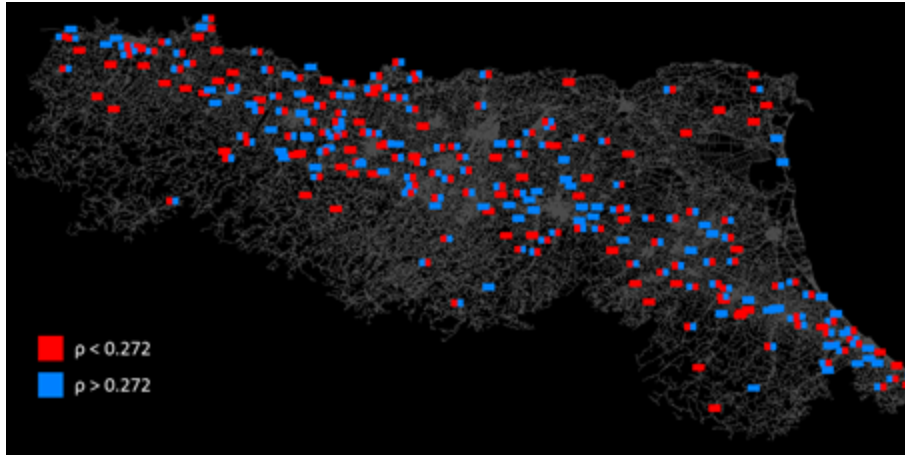


Figura 5.4: è mostrato il risultato su cartina della divisione in due fasce

Anche con questa suddivisione non è possibile trovare una chiara distribuzione delle differenze tra le penetrazioni mattutine e pomeridiane.

Di seguito è mostrato l'istogramma delle occorrenze delle coppie, sottolineando di nuovo che non è presente la totalità dei dati considerati precedentemente poiché le spire che su una delle due corsie mancano dei valori porta all'esclusione della coppia di dati.

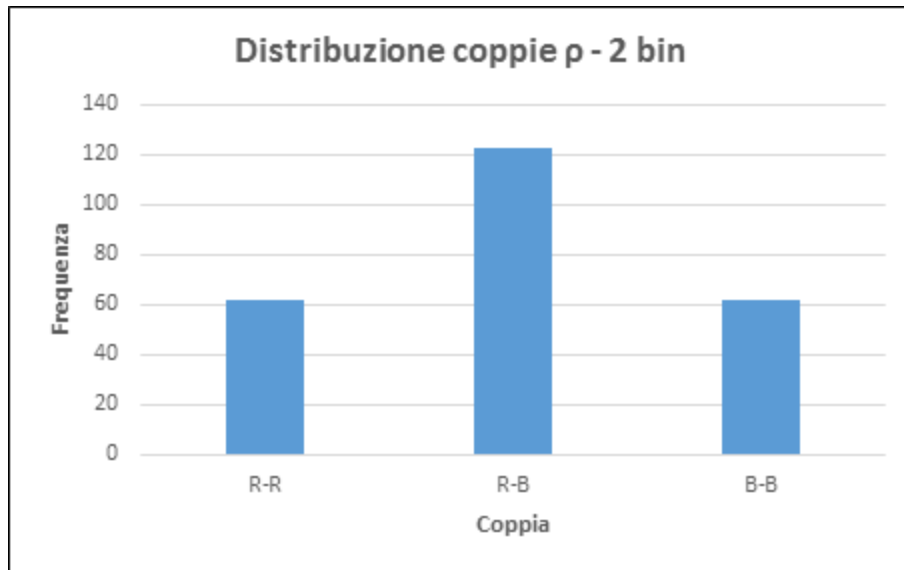


Figura 5.5: è mostrato l'istogramma delle occorrenze delle due fasce, ogni bin rappresenta una determinata coppia di range di valori.

È possibile notare che le occorrenze totali delle coppie omogenee, cioè con valori di penetrazione all'interno dello stesso range (R-R più B-B), sono circa uguali alle occorrenze in cui la coppia è eterogenea.

5.2.3 Conclusioni

Tutti questi risultati suggeriscono che il rapporto ρ , cioè la variazione di penetrazione fra mattina e pomeriggio, è una grandezza che fluttua molto ed assume valori, in percentuale, dallo 0 al 60%.

Inoltre, in generale le due direzioni di transito delle spire non sono correlate; non si hanno abbastanza informazioni, quindi, per prevedere la variazione di penetrazione di una corsia ignota a partire dalla sua opposta, poiché dall'osservazione empirica le due corsie fluttuano indipendentemente, e si hanno una equa distribuzione dei casi in cui una strada ha penetrazione costante lungo la giornata in un verso, e variabile nell'altro.

5.3 Probabilità di svolta

La seconda analisi che si è effettuata sulle fluttuazioni tra mattina e pomeriggio è stata quella riguardo alle probabilità di svolta, cioè la probabilità di imboccare

la stessa strada al variare delle due fasce della giornata. Si noti che proseguire sulla stessa strada invece che imboccarne un'altra ad un incrocio è considerata, ai fini delle analisi e del modello, alla stregua di una svolta.

Inizialmente ci si è basati esclusivamente sui dati dalle spire per il calcolo delle probabilità di svolta. Tuttavia, a causa dell'organizzazione della cartografia che si ha a disposizione, la grande maggioranza delle strade è divisa in segmenti che hanno identificativi diversi; questo può falsare il calcolo delle probabilità di svolta poiché se non si tiene conto di questa divisione in segmenti, ci saranno molte probabilità vicine al 100% fra i segmenti di ogni singola strada. Per motivi analoghi, le basse percentuali, che spesso sono sotto allo 0.5%, indicano veicoli che si sono fermati o spostati in una strada privata.

Per questo, quando si sono osservate solo le spire, la gran parte dei dati erano poco significativi e solo circa un centesimo delle spire avevano una statistica utile.

Quindi si è proceduto ad estendere il calcolo su tutta la rete regionale, selezionando le strade con una percorrenza totale nei 20 giorni considerati maggiore di 1000 auto dotate di GPS. Sono state escluse le strade dove la probabilità di svolta più alta risultava maggiore del 90% per ovviare al problema della segmentazione delle strade, nonché i casi in cui le svolte fossero molto basse.

Si è preferito lavorare con numeri più grandi e con statistiche più robuste. Per avere un'idea del valore di flusso medio per 1000 auto in 20 giorni con una penetrazione pari al 2%, sarebbe un flusso medio poco maggiore di 100 macchine all'ora.

Questa selezione ha portato ad avere una statistica di circa 12000 svolte.

L'istogramma che segue è organizzato nella seguente maniera: per ogni set di svolte uscenti da una strada, viene considerata solamente quella con probabilità maggiore; vengono escluse le strade in cui le svolte più percorse abbiano percentuali sotto al 50% perché il volume della statistica per quei dati peggiora molto, mentre come già accennato vengono ignorate svolte in cui la maggiore sia oltre al 90%.

Vengono quindi graficate le occorrenze per ogni range di svolta.

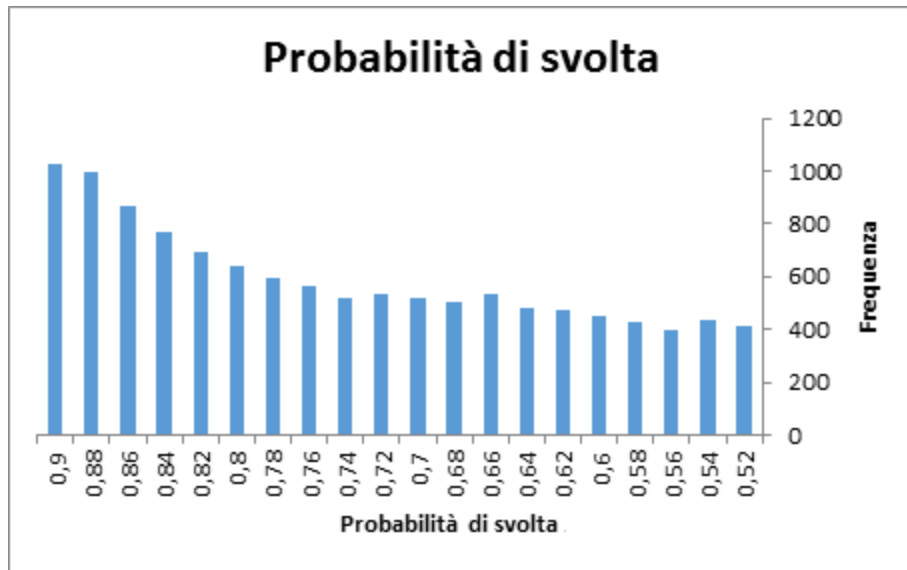


Figura 5.6: è mostrato l'istogramma delle occorrenze delle probabilità della svolta con percentuale maggiore per le strade selezionate.

Si è scelto di osservare di quanto le probabilità di svolta cambiassero tra mattina e pomeriggio in un modo simile a quello osservato per le penetrazioni, utilizzando a denominatore il valore delle svolte per l'intera giornata, adottando quindi la seguente formula:

$$\sigma_i = \frac{|p_svolta_{im}^{(matt)} - p_svolta_{im}^{(pom)}|}{p_svolta_{im}^{(giornata)}} \quad (5.2)$$

Si consideri che, come già accennato, ogni strada è identificata tramite la svolta con probabilità maggiore, ignorando le altre. Per la variazione mattina-pomeriggio si ottiene il seguente grafico:

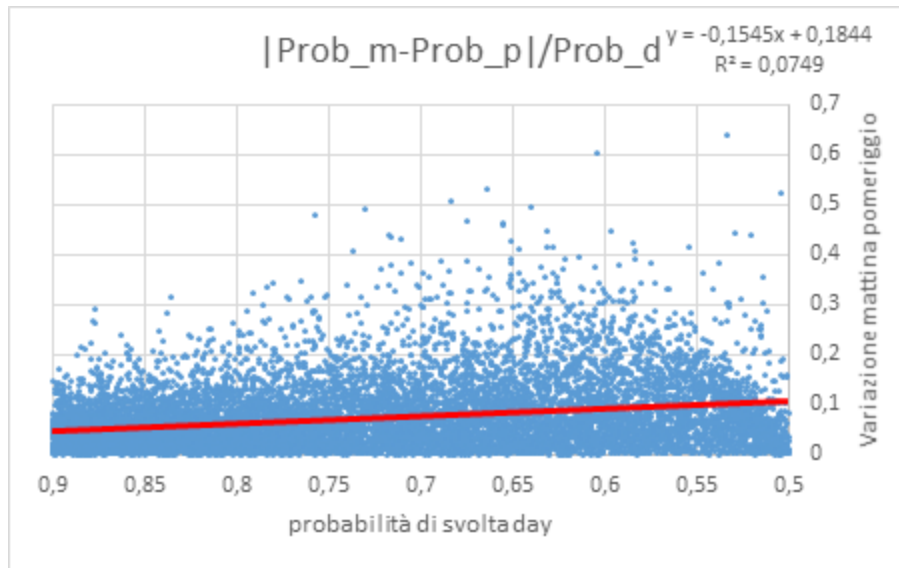


Figura 5.7: sono mostrate tutte le variazioni delle probabilità di svolta ordinate secondo il loro valore di svolta giornaliero, sempre considerando per ogni strada la svolta più probabile delle selezionate.

Si può riscontrare come per svolte con probabilità molto alta (sulla sinistra nel grafico) i punti siano abbastanza compatti, ed invece al diminuire della probabilità di svolta gli andamenti tra mattina e pomeriggio tendano a variare maggiormente.

Inoltre poiché sono presenti molti più dati per svolte con probabilità maggiori, se questa variazione fosse solo statistica si dovrebbe osservare la tendenza opposta. Una possibile spiegazione di questo sta nel fatto che nei casi in cui la probabilità di svolta è alta, la strada è scelta abbastanza univoca e quindi è relativamente improbabile che si adottino le altre svolte; invece quando le svolte da una strada diventano circa equiprobabili c'è più varianza nella scelta.

Si è quindi condotto un ulteriore studio sulla statistica di questi dati. Si sono divisi i dati in 20 bin, calcolando per ogni bin la media e varianza dei dati che contiene. Nel grafico seguente sono mostrati i valori interni ad una deviazione standard in blu e gli outliers della distribuzione: in verde i dati distanti più di una deviazione standard dal valor medio del bin, in giallo quelli distanti più di due e in rosso quelli distanti più di tre.

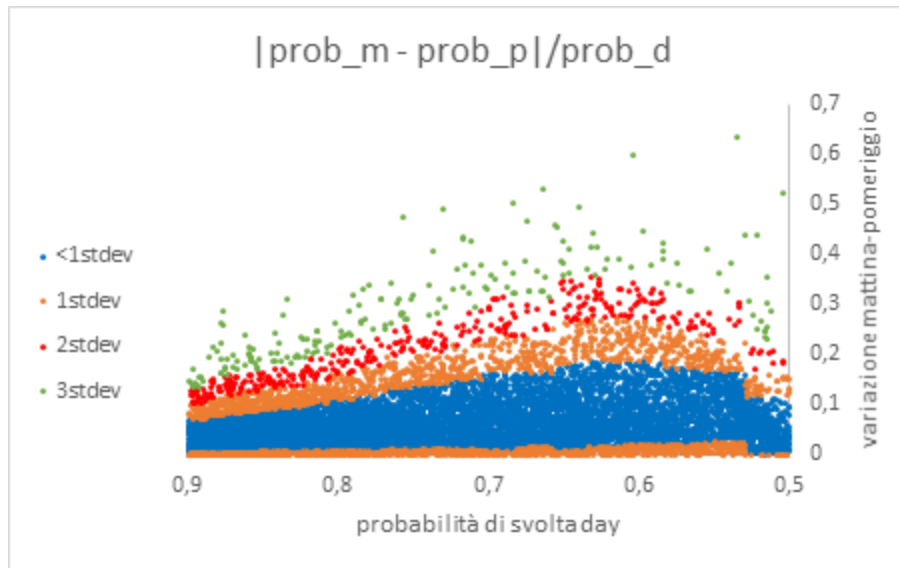


Figura 5.8: sono mostrate tutte le variazioni delle probabilità di svolta ordinate secondo il loro valore di svolta giornaliero, colorate in base alla distanza dal valor medio.

Raccolte le informazioni sugli outliers, si sono mostrate sulla mappa per osservare le zone con alta variazione giornaliera. Nelle immagini successive sono mostrate alcune province della regione. Le strade colorate rappresentano gli outliers, mentre quelle in grigio sono nella media o sono tratti con probabilità di svolta del 100%, cioè segmenti interni di strade poligonali.

I colori delle strade mostrate nella mappa seguono la seguente legenda cromatica:

- in grigio scuro sono le strade con basso flusso che sono state escluse da tutti i calcoli del programma;
- in grigio chiaro sono le strade in cui la variazione della probabilità di svolta è minore del 10% o il flusso è inferiore alla soglia sopracitata dei 1000 veicoli totali nei 20 giorni;
- in blu sono rappresentate le strade che rientrano all'interno di una deviazione standard dal valor medio del proprio bin;
- in arancione quelle entro due deviazione standard;

- in rosso entro tre deviazioni standard;
- infine in verde quelle distanti oltre tre deviazioni standard.

Si è scelto di evitare di mostrare l'intera mappa della regione perché, a causa della risoluzione che avrebbe l'immagine, sarebbe di impossibile lettura; di seguito è mostrata quindi la mappa delle svolte nella città di Bologna ed i suoi dintorni:



Figura 5.9: è inserita la mappa di Bologna con le svolte prese in considerazione colorate secondo la scala corantica descritta precedentemente.

Modena:



Figura 5.10: è inserita la mappa di Modena con le svolte prese in considerazione colorate secondo la scala corantica descritta precedentemente.

Rimini:

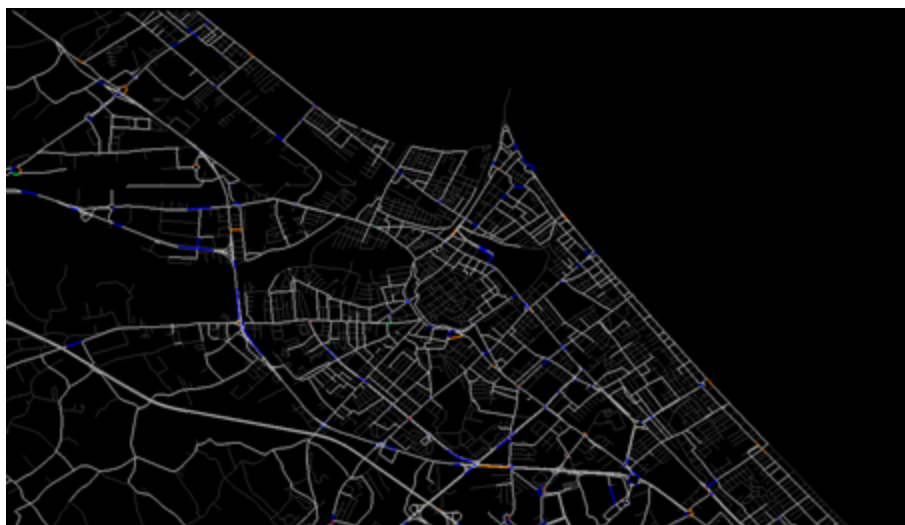


Figura 5.11: è inserita la mappa di Rimini con le svolte prese in considerazione colorate secondo la scala corantica descritta precedentemente.

Da queste immagini si vede che il numero degli outliers è molto basso e sparso. Comunque le variazioni percentuali restano sotto al 20% per tutti i dati, cioè sotto ad una deviazione standard, quindi si è concluso che non fosse rilevante introdurre questo cambiamento nel modello.

5.4 Validazione della simulazione

Il terzo parametro preso in esame è stato quello della validazione del modello, cioè si è studiato se l'indice di bontà della simulazione migliorasse nel considerare i diversi transiti tra mattina e pomeriggio.

Il metodo di validazione, come già spiegato nel “Capitolo 3”, si basa sulla simulazione dell'intero modello cancellando di volta in volta una spira e sull'osservazione di come il modello riesca ad attribuire un valore a questa; ciò porta ad ottenere una stima per eccesso dell'errore commesso tra le strade comprese fra una spira e le sue prime vicine. Il valore di bontà della simulazione è il valor medio di questi errori.

I parametri utilizzati in questa parte di simulazione sono il numero di auto che transitano su una strada ed il valore di penetrazione di quella strada.

Il numero totale di auto è stato calcolato a partire dai dati di transito della strumentazione GPS suddividendoli in due fasce, quella mattutina e quella pomeridiana.

La ricalibrazione delle penetrazioni è già stata spiegata nel paragrafo precedente.

Le formule che sono alla base della simulazione sono le seguenti:

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

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

dove negli intervalli orari relativi alla mattinata i valori $n_i^{(out)}$ e p_gps_i sono quelli relativi alla mattina ed analogamente nel pomeriggio. I valori di flusso vengono confrontati con il flusso effettivamente misurato dalle spire e dopo essere stati sommati vengono nuovamente confrontati quelli calcolati ignorando la

divisione mattina pomeriggio. Si è osservato un miglioramento del 7% sul calcolo dell'errore del sistema, e sono stati graficati diverse grandezze che permettono di osservare relazioni fra i valori.

Eliminando i casi anomali si è evidenziata una relazione tra l'errore totale giornaliero commesso sulle strade dotate di spira ed il flusso giornaliero medio.

Si può osservare il primo grafico dove vengono mostrati tutti gli errori in funzione dell'intensità del flusso, mentre nel secondo è presente l'errore relativo. I grafici sono riportati prima utilizzando i valori senza la divisione in fasce orarie mentre successivamente è presente la divisione tra mattina e pomeriggio.

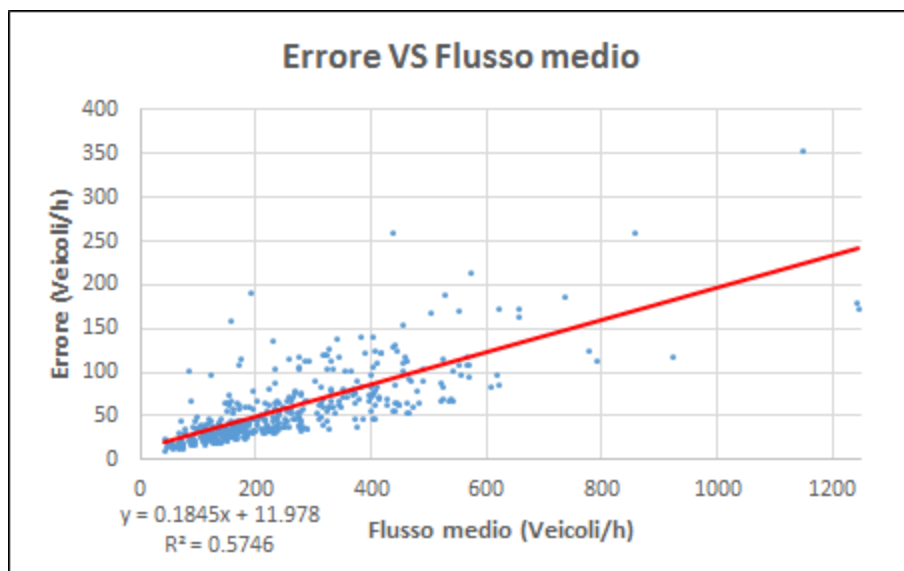


Figura 5.12: è rappresentato in un grafico a dispersione l'errore assoluto commesso in ordinata mentre in ascissa il valore di flusso medio giornaliero.

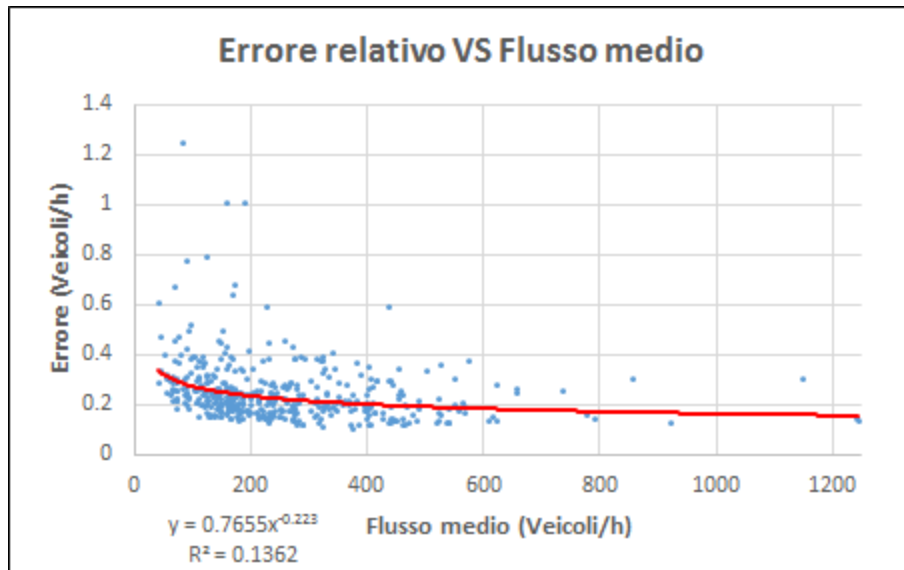


Figura 5.13: è rappresentato in un grafico a dispersione l'errore relativo commesso in ordinata mentre in ascissa il valore di flusso medio giornaliero.

Introducendo la divisione nelle due fasce giornaliere:

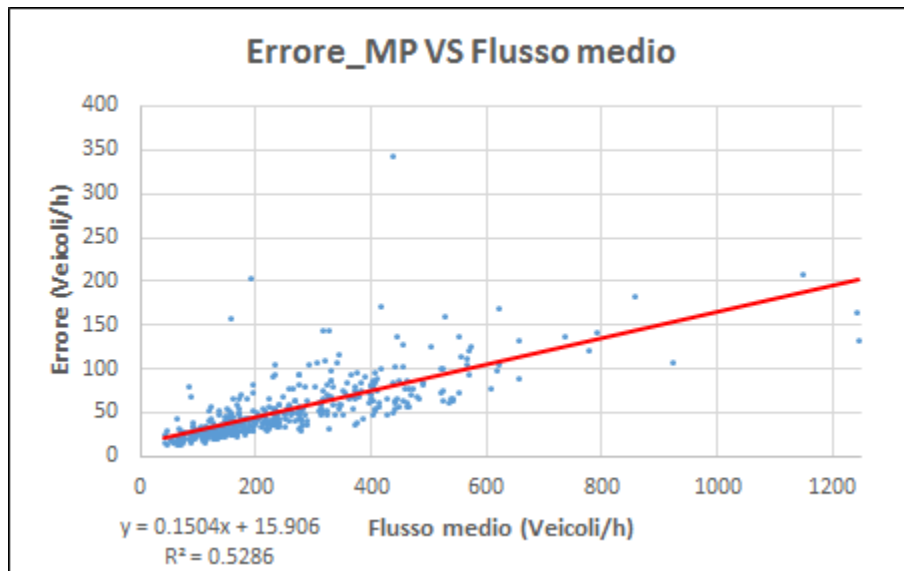


Figura 5.14: è rappresentato in un grafico a dispersione l'errore assoluto commesso in ordinata mentre in ascissa il valore di flusso medio giornaliero.

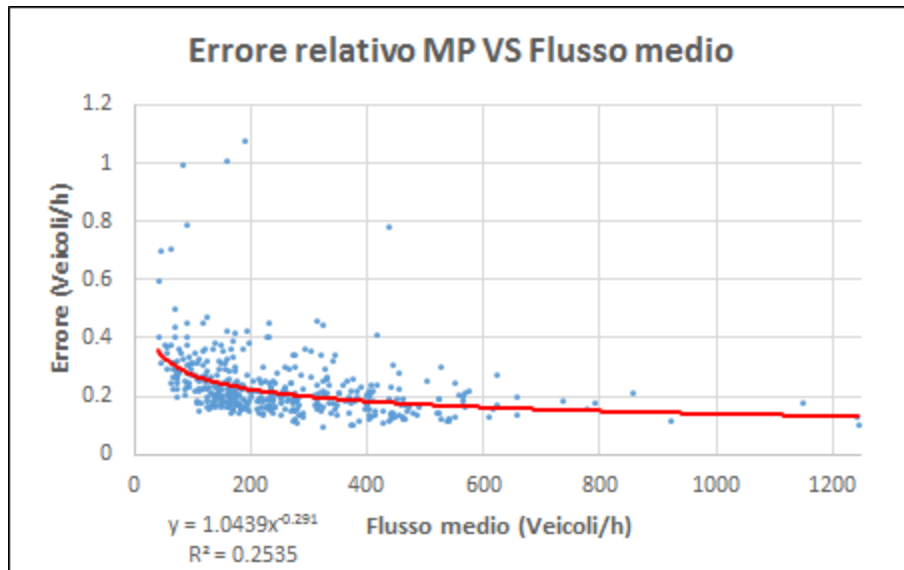


Figura 5.15: è rappresentato in un grafico a dispersione l'errore relativo commesso in ordinata mentre in ascissa il valore di flusso medio giornaliero.

Si ripropone qui un grafico in cui appare per ogni spira l'errore, in questo caso quelli ottenuto con la divisione in fasce orarie, ordinando le spire in base alle distanze di correlazione medie delle proprie influenti.

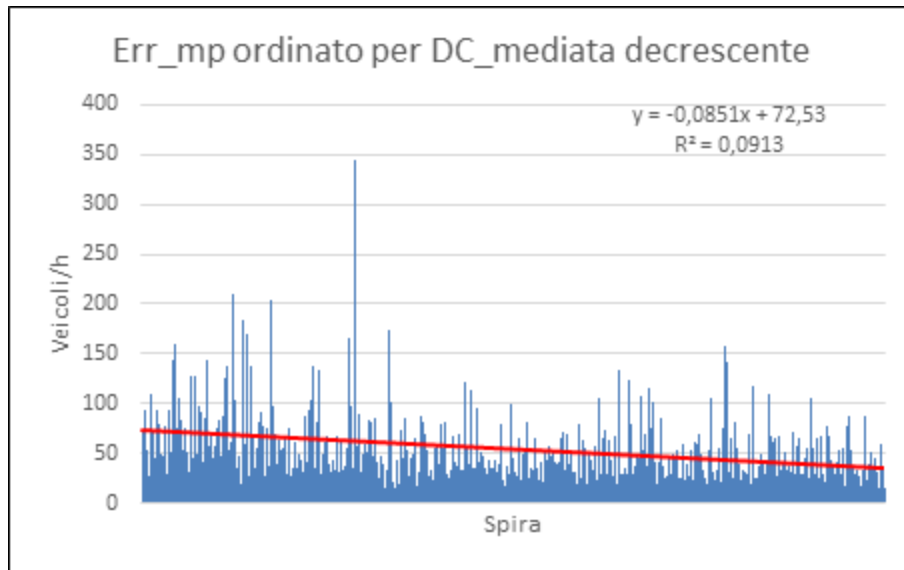


Figura 5.16: è rappresentato l'errore commesso per ogni spira ordinandole per valore di distanza di correlazione decrescente.

Confrontando questo grafico con quello presentato nel "Capitolo 4", si nota che la differenza principale è che il valor medio degli errori è minore rispetto al caso in cui non si divide in fasce la giornata. Ma soprattutto il risultato del confronto del parametro di validazione mostra un miglioramento del 7% nei valori degli errori, si veda il "Capitolo 6" per approfondire questo aspetto.

Capitolo 6

Risultati del modello

6.1 introduzione

Durante lo svolgimento di questo lavoro di tesi, le varie ipotesi e sperimentazioni hanno portato a modificare il modello ed i suoi parametri diverse volte, ed in questo capitolo vengono discussi i risultati delle simulazioni. Come già esposto nel “Capitolo 3”, 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

$$E = \sum_i \int_{t_0}^T \epsilon_i(t) dt \quad (6.1)$$

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.

$$\bar{E} = \frac{E}{96 N} \quad (6.2)$$

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%. Per questa ragione gli errori mostrati in questo capitolo, ove non altrimenti specificato, saranno quelli che risultano dall'attuazione di questa suddivisione.

Con il 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, in quanto questa grandezza caratteristica si definisce come l'85° percentile del range di velocità misurato in ogni strada.

6.2 Risultati delle varie combinazione di pesi e metodi di propagazione

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

Utilizzando questi parametri, i pesi sono calcolati come

$$w_{ji} = \frac{1}{t_trip_{ij}} \quad (6.3)$$

E la propagazione avviene con

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p_gps_i}{n_j^{(out)}/p_gps_j} \varphi_j(t) \quad (6.4)$$

Il grafico rappresentato di seguito mostra l'errore medio delle spire rappresentato in un istogramma che mostra la sua distribuzione.

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

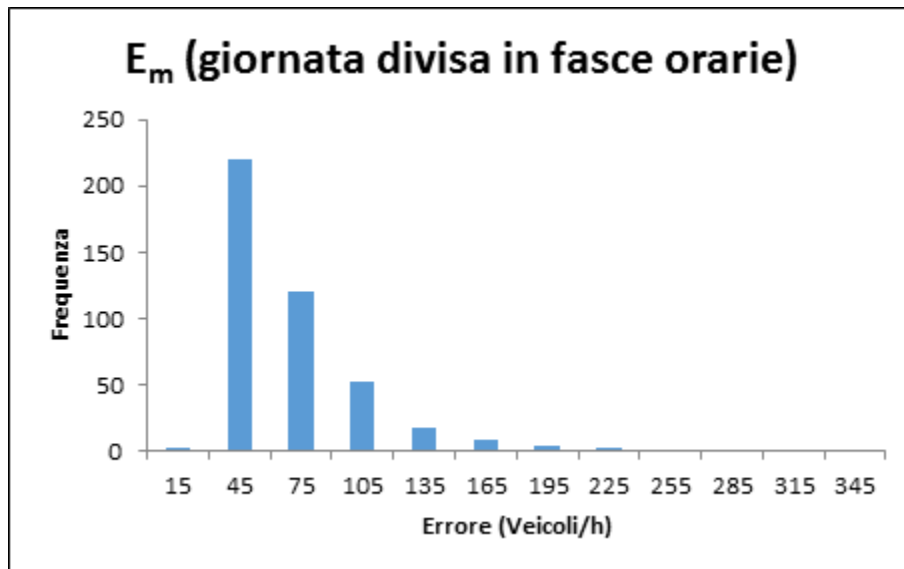


Figura 6.1: è mostrato l'istogramma degli errori commessi per ogni spira utilizzando il metodo sopracitato. Si noti che nell'algoritmo sono utilizzati i valori di penetrazione e di flussi relativi alle fasce orarie quali mattina e pomeriggio

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 \bar{E} vale 54.2 veicoli/h.

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

Di seguito viene mostrato l'errore rappresentato cromaticamente sulla mappa regionale, per dare un'idea delle zone rappresentate 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 nel "Capitolo 3", sarà minore del valore espresso.

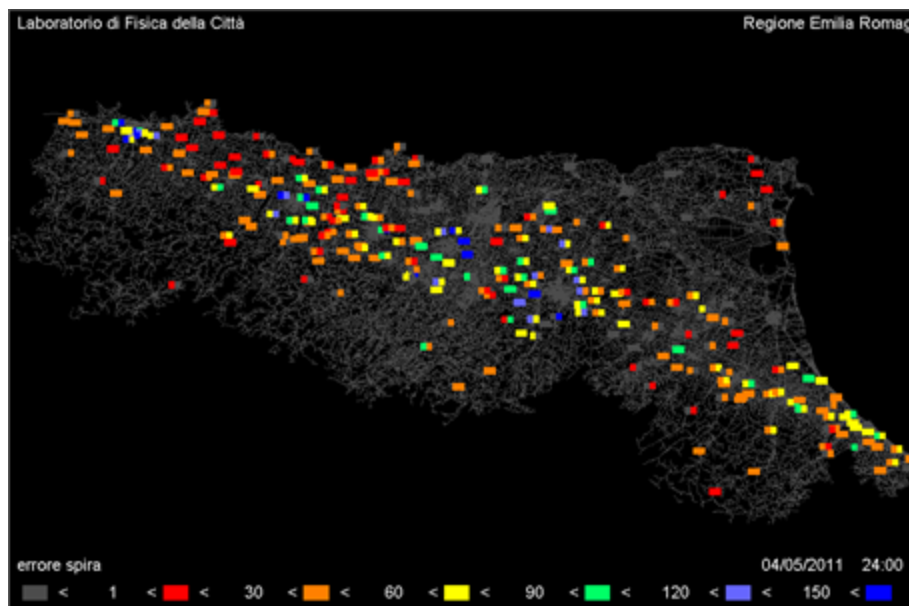


Figura 6.2: è mostrato il valore di errore assoluto su ogni postazione di spira in mappa.

Si noti che si è adottata la convenzione di rappresentare ogni postazione di 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; Questo è dovuto dal fatto che attorno ai centri urbani e lungo la via Emilia il flusso di auto è alto e che l'errore rappresentato è assoluto.

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

In questo caso i pesi sono calcolati come:

$$w_{ji} = \frac{1}{t_trip_{ij}} \quad (6.5)$$

e la propagazione avviene con

$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t) \quad (6.6)$$

Analogamente al paragrafo precedente, si propone il grafico della distribuzione dell'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, e come si è già osservato le percentuali di svolta non variano in maniera significativa per essere inserite.

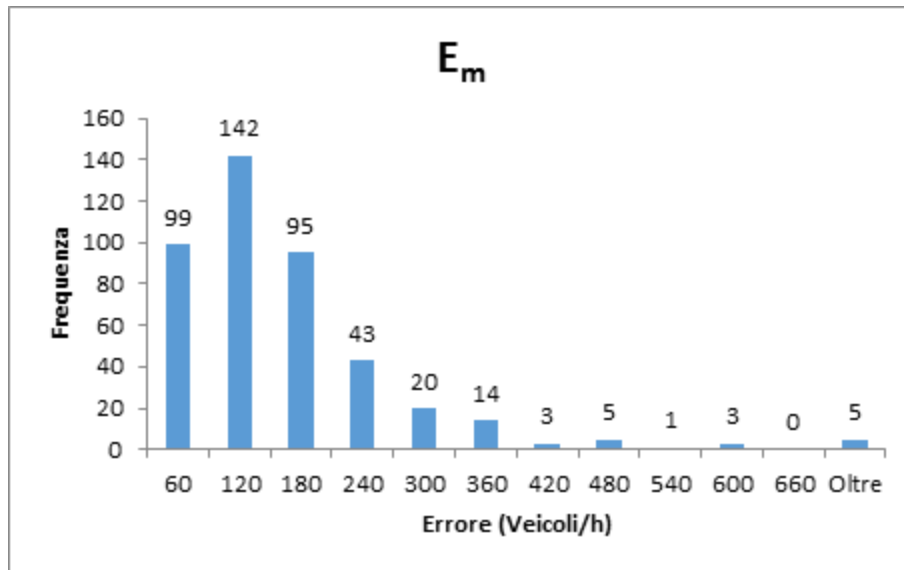


Figura 6.3: è mostrato l'istogramma degli errori commessi per ogni spira utilizzando il metodo sopracitato.

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 \bar{E} è 138.7 veicoli/h.

È però evidente una distribuzione statistica simile a prima.

Di seguito è mostrato l'errore sulla mappa.

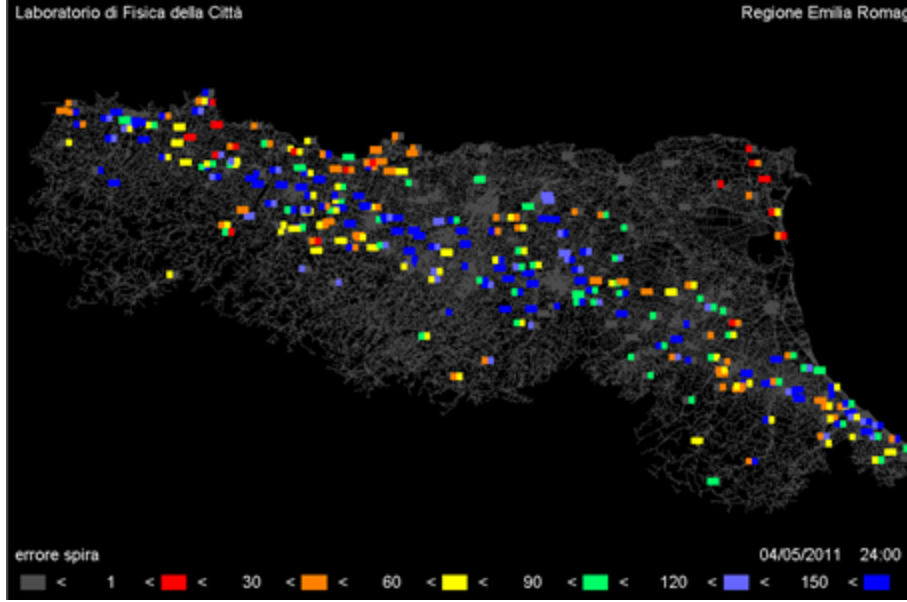


Figura 6.4: è mostrato il valore di errore assoluto su ogni postazione di spira in mappa.

Questo algoritmo di propagazione non è altrettanto buono rispetto al precedente, e anche al di fuori dei centri urbani si vedono aree con errore elevato e questo non è bilanciato dall'ampio flusso in quelle zone rispetto a quelle cittadine.

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

In questo terzo schema, i pesi sono definiti come

$$w_{ji} = p_svolta_{ij} = \frac{n_j^{(out)}}{n_i^{(out)}} \quad (6.7)$$

E la propagazione, esattamente come nel primo metodo:

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p_gps_i}{n_j^{(out)}/p_gps_j} \varphi_j(t) \quad (6.8)$$

Questo metodo dà risultati poco diversi dal primo, poiché il sistema dipende più debolmente dalla scelta dei pesi che dal metodo di propagazione; il grafico che rappresenta la distribuzione dell'errore non mostra troppe differenze rispetto al primo.

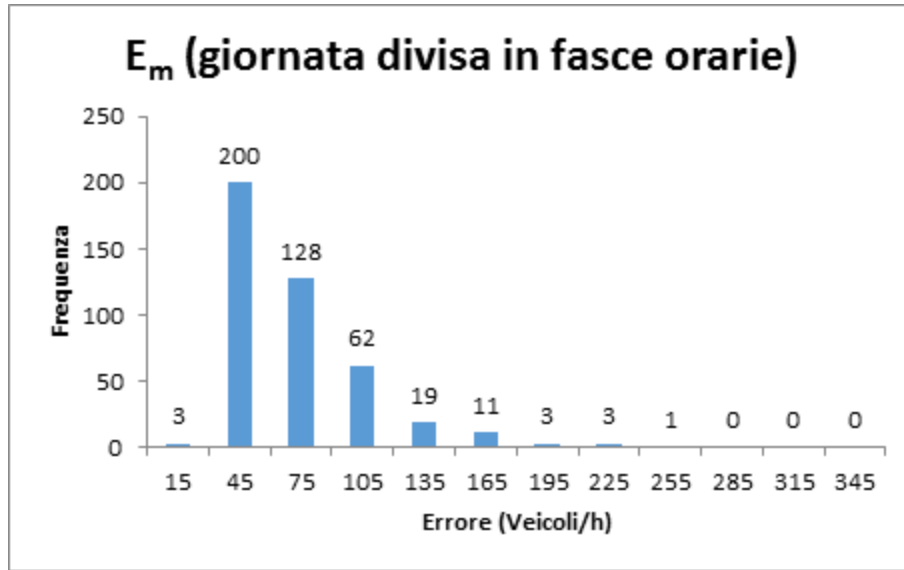


Figura 6.5: è mostrato l'istogramma degli errori commessi per ogni spira utilizzando il metodo sopracitato. Si noti che nell'algoritmo sono utilizzati i valori di penetrazione e di flussi relativi alle fasce orarie quali mattina e pomeriggio

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.

6.2.4 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

$$w_{ji} = p_svolta_{ij} = \frac{n_j^{(out)}}{n_i^{(out)}} \quad (6.9)$$

E la propagazione avviene con

$$\varphi_i(t) = \sum_j w_{ji} \varphi_j(t) \quad (6.10)$$

Segue il grafico che mostra la distribuzione dell'errore:

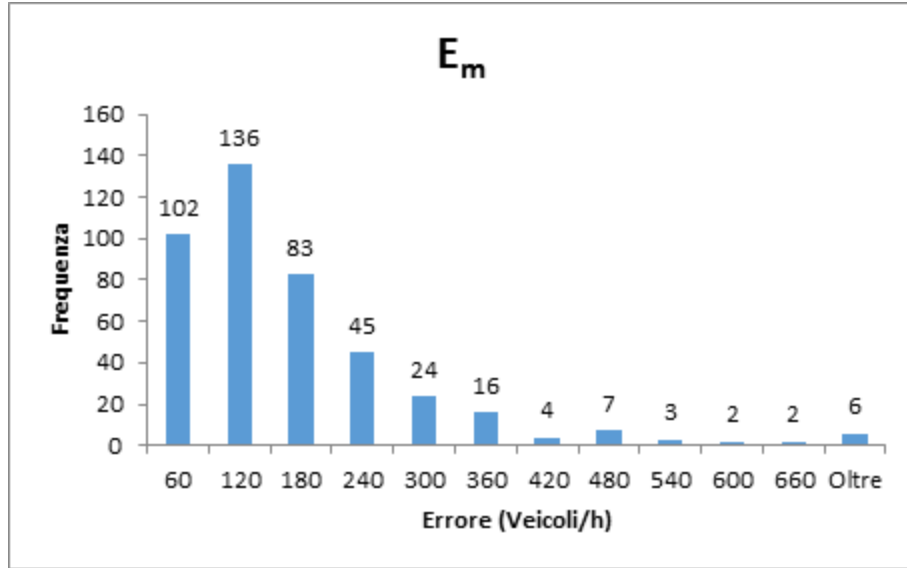


Figura 6.6: è mostrato l'istogramma degli errori commessi per ogni spira utilizzando il metodo sopracitato.

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.

6.3 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

$$\varphi_i(t) = \sum_j w_{ji} \frac{n_i^{(out)}/p_gps_i}{n_j^{(out)}/p_gps_j} \varphi_j(t) \quad (6.11)$$

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(Auto/h)	Errore MP(Auto/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	-

Tabella 6.1: è mostrata la tabella che riassume le varie combinazioni dei metodi e dove possibile anche i valori di errore calcolati utilizzando la divisione tra mattina e pomeriggio.

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% e si è confrontato il risultato del modello migliore (tempi di propagazione e rapporti di flussi) con lo stesso ma con penetrazione fissata al 2% e si è ottenuto questo risultato:

Penetrazione	Errore (Auto/h)	Errore MP (Auto/h)
Empirica	58.9	54.2
Fissa a 2%	79.9	-

Tabella 6.2: è mostrata la tabella che mostra il miglioramento utilizzando l'algoritmo migliore con i valori di penetrazione variabili da quelli fissati a 2%.

Volendo schematizzare la validità dei vari algoritmi, si propone il seguente diagramma.

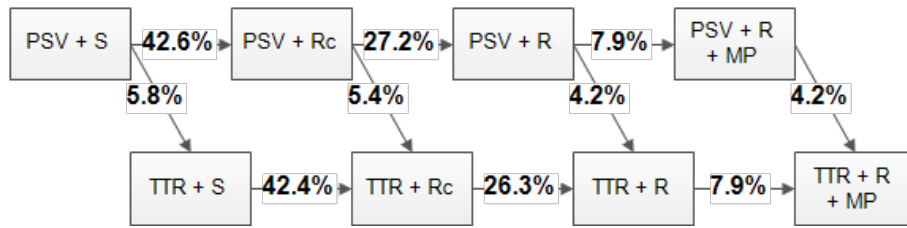


Figura 6.7: è mostrato un grafico che mostra i miglioramenti percentuali tra i differenti metodi nel calcolare l'errore medio secondo la legenda sottostante.

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.

Nella grafico successivo si può osservare un confronto diretto degli errori dei quattro tipi di algoritmo per il modello.

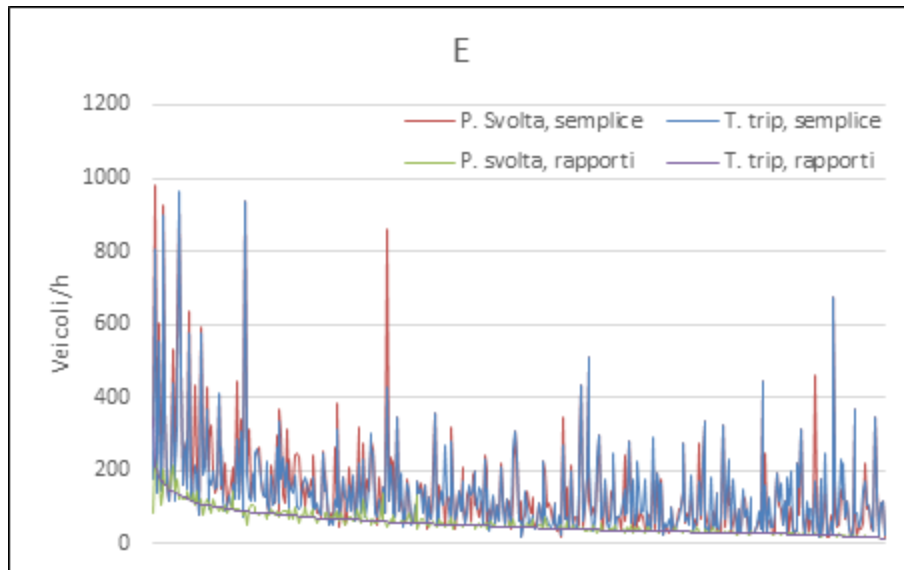


Figura 6.8: sono mostrati i valori degli errori per spira ordinati secondo il valore di errore decrescente del metodo che meglio riduce l'errore globale.

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 sua assenza negli altri.

Si evince, inoltre, che i tre metodi inferiori riproducono meglio i flussi solo in pochi punti, ed in misura minima, quindi anche componendo un metodo misto prendendo per ogni spira il metodo migliore di propagazione non si otterrebbe un errore molto minore.

6.4 Calcolo del campo di flusso nella regione

Una volta osservati i risultati dei differenti metodi, è opportuno osservare 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, che come già mostrato in precedenza è 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. Si osserva che c'è coerenza spaziale nei flussi ricostruiti e gli errori sono sufficientemente limitati per poter affermare che l'estensione del campo è avvenuta con relativo successo.

Ore 1:00

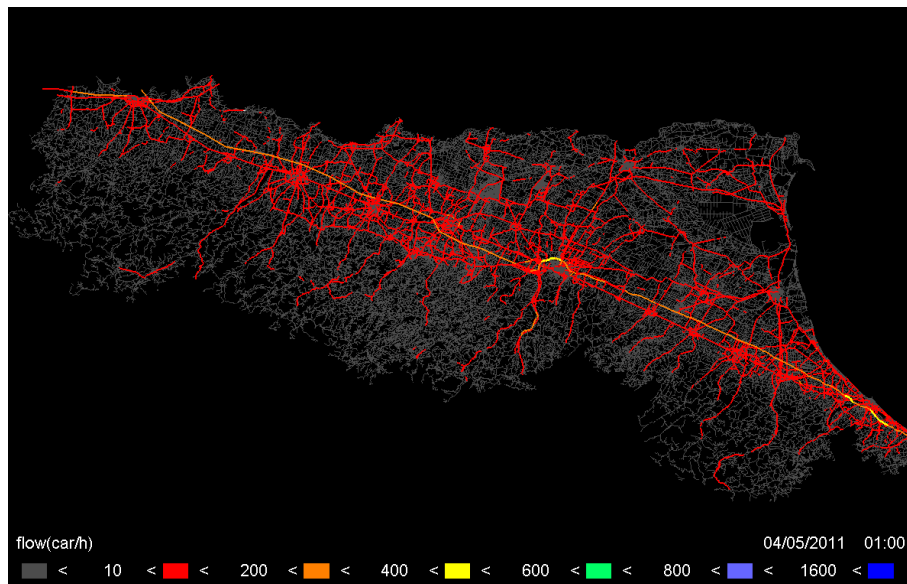


Figura 6.9: è rappresentata la ricostruzione di flusso alle ore 1:00.

Si osserva come i flussi medi sono molto bassi tranne alcuni flussi importanti 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.

Ore 4:00

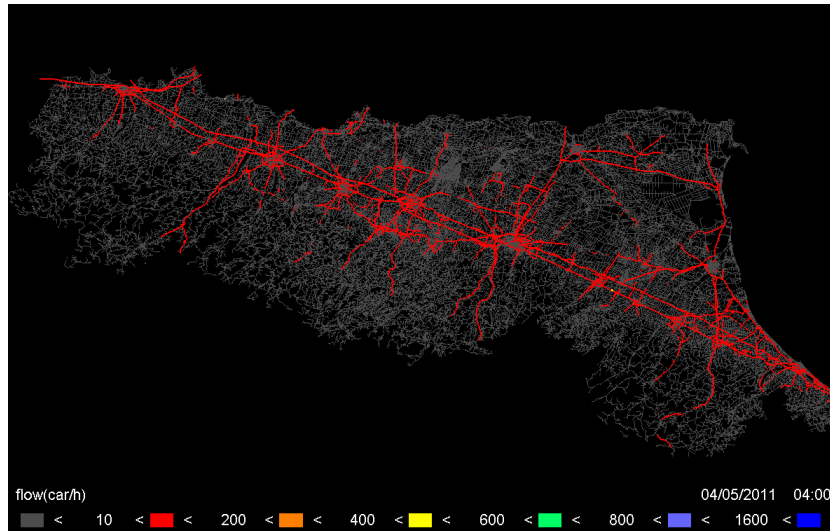


Figura 6.10: è rappresentata la ricostruzione di flusso alle ore 4:00.

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

Ore 6:00

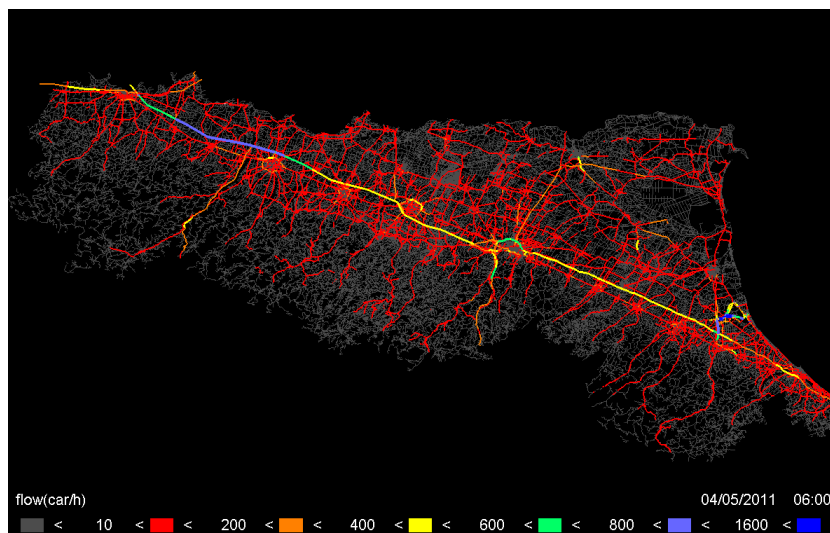


Figura 6.11: è rappresentata la ricostruzione di flusso alle ore 6:00.

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

Ore 8:00

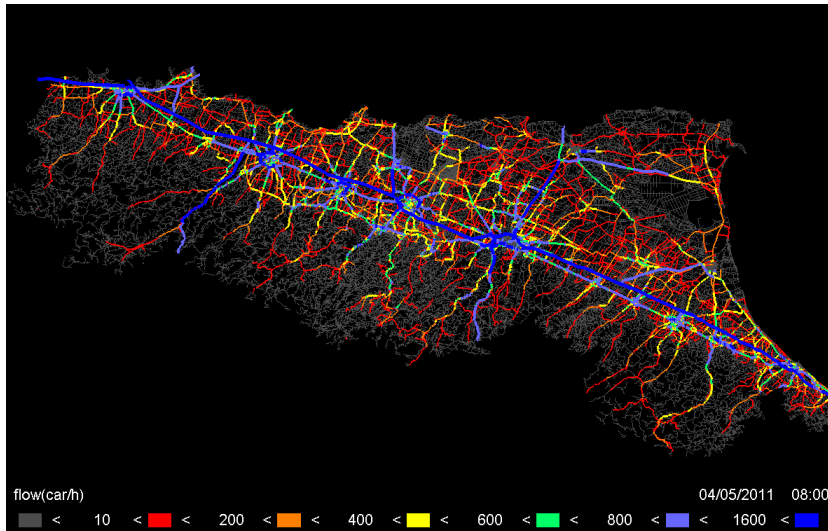


Figura 6.12: è rappresentata la ricostruzione di flusso alle ore 8:00.

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.

Ore 10:00

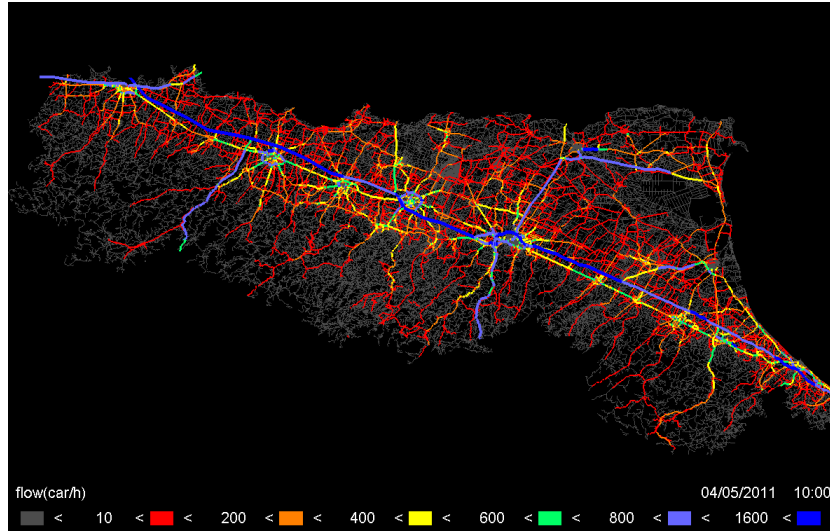


Figura 6.13: è rappresentata la ricostruzione di flusso alle ore 10:00.

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.

Ore 18:00

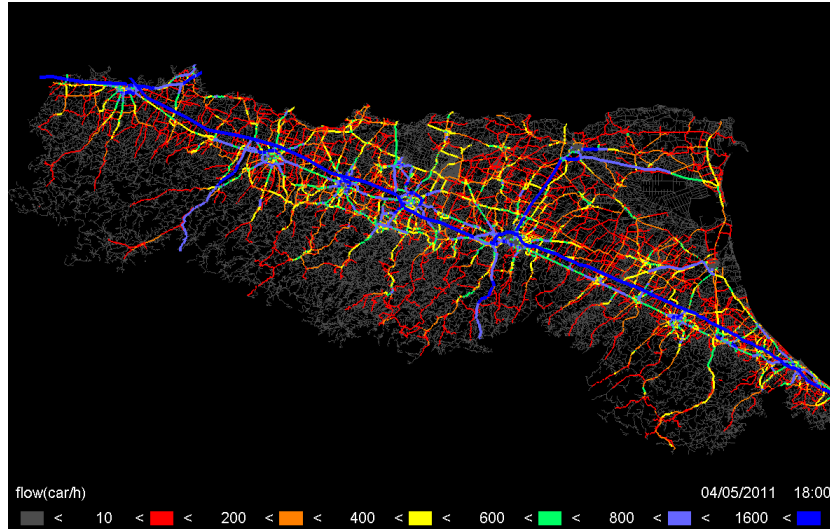


Figura 6.14: è rappresentata la ricostruzione di flusso alle ore 18:00.

Si osserva il secondo massimo della giornata dove la regione ha mediamente un flusso simile a quello del picco delle 8:00.

Ore 20:00

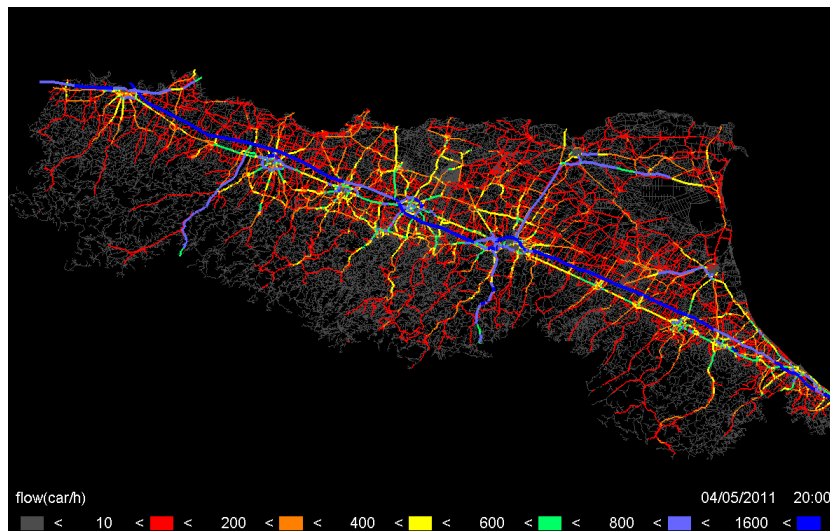


Figura 6.15: è rappresentata la ricostruzione di flusso alle ore 20:00.

Si nota che i flussi iniziano a diminuire, trend che continuerà fino a circa le 4:00.

Ore 23:00

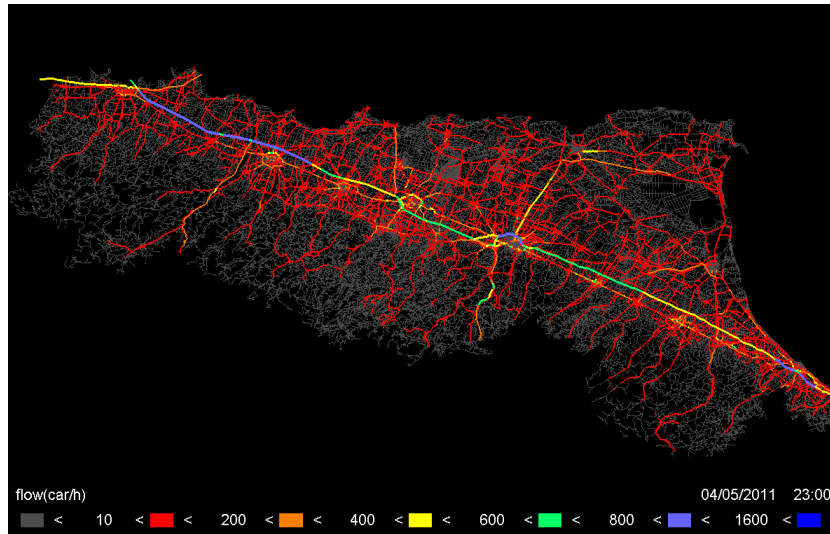


Figura 6.16: è rappresentata la ricostruzione di flusso alle ore 23:00.

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

6.5 Free Speed

Per quanto riguarda la free speed, le figure non variano nel tempo poiché i valori sono medie calcolate a partire da tutte le velocità misurate nel mese, attraverso ai dati GPS selezionando i valori del 85 esimo percentile.

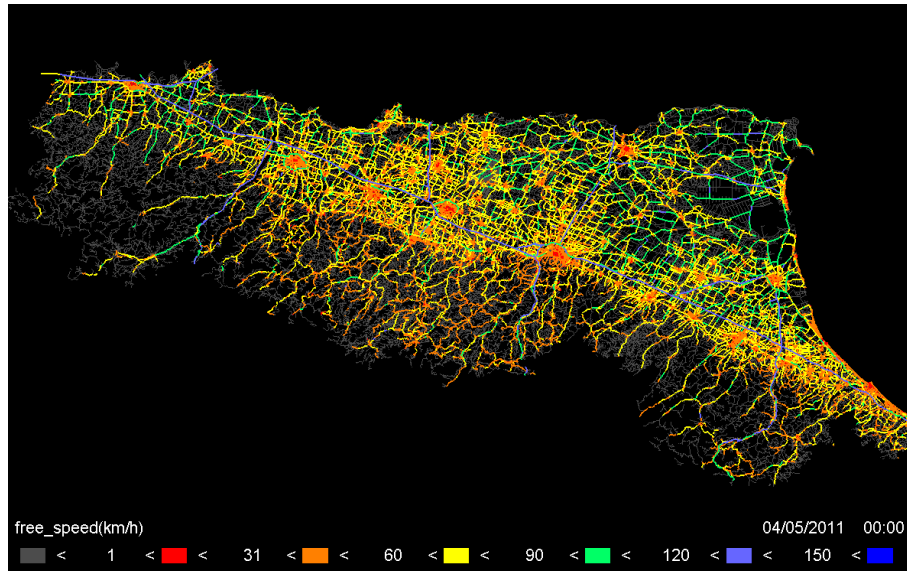


Figura 6.17: sono rappresentate le free speed nella regione calcolate dai dati GPS.

Come era plausibile prevedere all'interno delle città sono presenti diverse macchie rosse, in cui le velocità sono molto basse, poi più ci si sposta dai centri urbani più le velocità aumentano, in arancione nelle zone ancora vicine alle città medio-grandi o nelle zone degli appennini mentre quasi la totalità delle strade extra urbane sono in giallo. Le strade in verde indicano scorrimenti molto veloci e spesso sono alternate a tratti di violetto che maggiormente si trovano su tratti autostradali.

Capitolo 7

Conclusioni

L'intento di questo lavoro di tesi era quello di sviluppare un modello in grado di ricostruire i flussi nell'intera regione Emilia Romagna utilizzando due dati di natura differente e possedendo una cartografia di fondo, questo obiettivo si può considerare conseguito.

Il lavoro fondamentale di questa tesi è quello di inserire nel modello le due fonti di informazioni che per loro natura sono molto differenti, infatti i dati da spira sono considerati precisi senza errore, ma molto rarefatti nello spazio, mentre i dati da GPS sono al contrario molto diffusi, per quanto siano distribuiti in maniera disomogenea nello spazio, e gli apparati sono stati utilizzati da un gruppo di popolazione anch'esso disomogeneo.

È chiaro che se i dati fossero perfetti, ossia una penetrazione del 100% o se ogni strada fosse dotata di spire, uno solo dei due dati permetterebbe di ricostruire il flusso senza alcun tipo di errore. Quindi partendo da questa banale assunzione si può pensare come migliorare il modello: infatti con i dati a disposizione al momento sono state esplorate tutte le possibilità che il gruppo di lavoro ha finora proposto e si pensa che per limitare ancor di più l'errore, che ha già valori molto interessanti, l'unico modo sia quello di aumentare il numero di fonti di dati, cioè aumentare la penetrazione dei dispositivi GPS o l'inserimento di apparati di spire dove si vuole ridurre l'errore.

È da sottolineare che al momento diverse compagnie di assicurazione e diverse compagnie private si stanno interessando ad inserire rivelatori GPS sempre in

più autoveicoli, quindi si può pensare che il modello stia diminuendo il suo valore di errore autonomamente più i dati che vengono utilizzati sono recenti.

Si valuti che già molte regioni sono dotate dei sistemi di acquisizione dati sovraccitati e quindi con poco sforzo sarebbe possibile esportare il programma ed ottenere gli stessi risultati in Lazio[2], Lombardia[7] e Piemonte.

È interessante evidenziare che il modello concluso sarà inserito nel sito della regione per mostrare i risultati della ricostruzione dei flussi real-time.

Bibliografia

- [1] M. Treiber and A. Kesting, "Traffic Flow Dynamics", Springer, 2013
- [2] M.G. Messina, G. Valenti, F. Carapellucci G. Fusco, C. Colombaroni, A. Gemma, G. Ciccarelli, S. Lo Sardo, ENEA, RICERCA DI SISTEMA ELETTRICO Sistema di monitoraggio e previsione della mobilità veicolare per l'integrazione tra la rete della illuminazione pubblica e la rete della mobilità, sapienza università di roma, 2011,p. 94
- [3] Dr. Ing., Sabato G. Rilevazione, conteggio e classificazione del traffico mediante spire induttive, 2006, http://www.doingsecurity.it/adon/files/mc2024t6_it_20040116_traffic.pdf
- [4] Gonzales Woods, Digital Image Processing 2nd Edition (DIP/2e), 2002
- [5] Dr, S. Rambaldi, "ProgEtto per la Gestione della mobilità Attraverso Sistemi infotelematici per l'ambito Urbano, per la Sicurezza di passeggeri, veicoli e merci (Pegasus)", Algoritmi di monitoraggio, previsione e controllo 2012
- [6] Annuario statistico 2011 ACI, <http://www.aci.it/>
- [7] F. Ferrari, "Flussi di traffico nel comune di Milano: dati rilevati ed analisi, Corso di laurea magistrale in Architettura Politecnico di Milano, 2011"