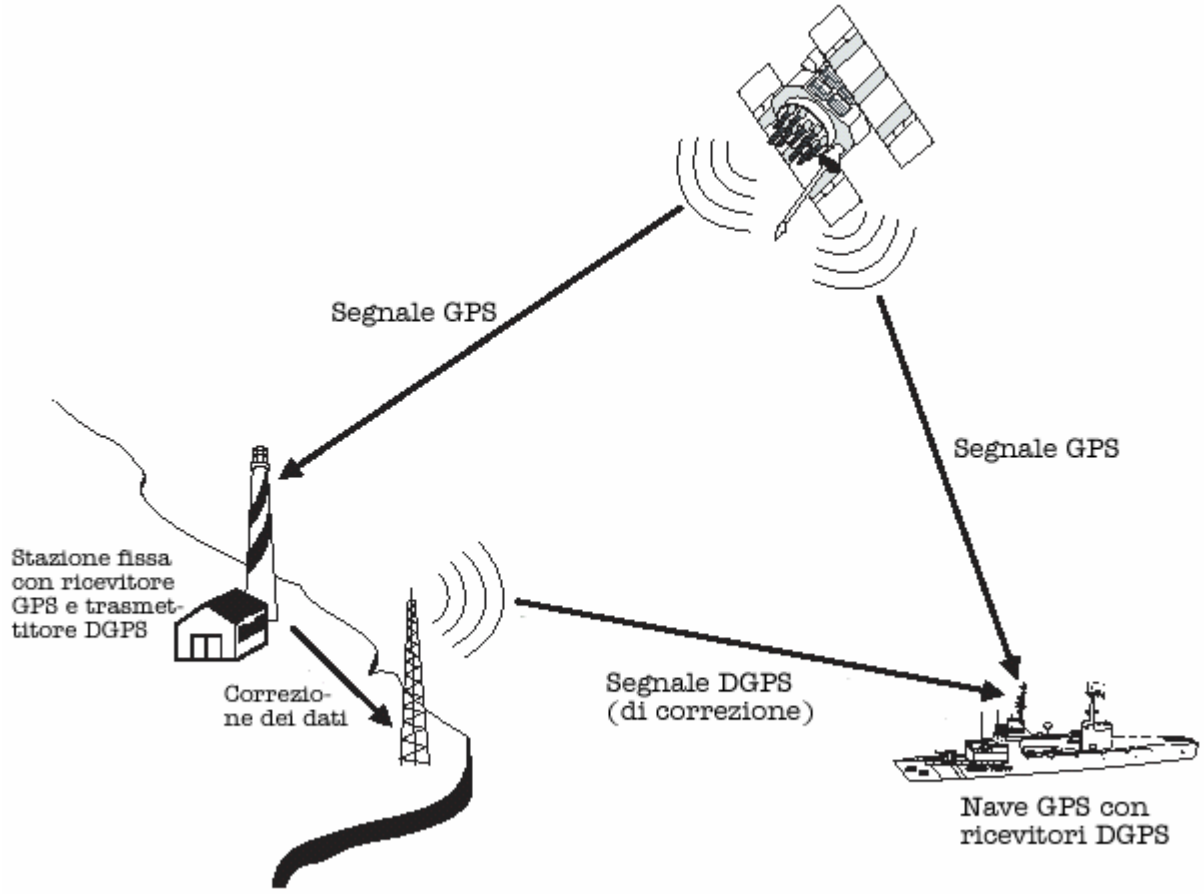
Capitolo 4

Il DGPS e gli algoritmi RTK/OTF

Fondamentalmente il GPS è il più accurato sistema di navigazione radio mai scoperto. E per molte applicazioni risulta essere molto accurato. Ma è nella natura umana pretendere di più. Così alcuni abili ingegneri hanno sviluppato il GPS differenziale, che è in sostanza un modo per correggere i vari errori del sistema GPS migliorando la precisione. Il DGPS (*Differential GPS*) può effettuare misurazioni precise fino ad un paio di metri per applicazioni in movimento e persino migliori in situazioni stazionarie.
Questo miglioramento della precisione ha un effetto profondo sull'importanza del GPS come risorsa. In questo modo esso diventa più di un sistema per guidare navi ed aeroplani lungo le loro rotte, esso diventa un sistema di misurazione universale capace di posizionamenti in una scala molto precisa.
Il DGPS necessita della cooperazione di due ricevitori, uno è stazionario e costituisce la cosiddetta "stazione base" e l'altro è in moto il cosiddetto "*rover*" ed effettua le misure di posizionamento. Un ruolo molto importante è giocato dal ricevitore stazionario, esso lega tutte le misure ricavate dai satelliti ad un sistema di riferimento locale tridimensionale.



Si ricordi che i ricevitori GPS usano i segnali cronometrici di almeno quattro satelliti per stabilire una posizione. Ognuno di questi segnali cronometrici è affetto da errori o ritardi dipendenti dal tipo di ostacoli che si sono trovati di fronte durante il viaggio. Fortunatamente la semplice scala del sistema GPS viene incontro agli utenti. I satelliti sono così distanti nello spazio che le piccole distanze che noi consideriamo sulla Terra sono insignificanti. Così se due ricevitori sono molto vicini, ad esempio meno di 100

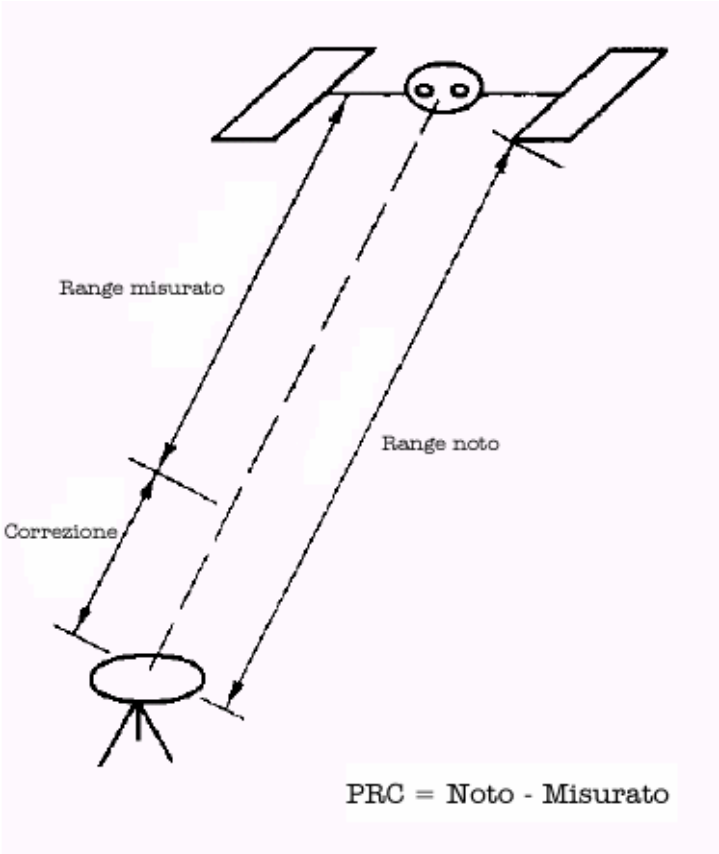
chilometri, i segnali che li raggiungono hanno viaggiato attraverso lo stesso "pezzo" di atmosfera e quindi sono virtualmente affetti dagli stessi errori.
Il DGPS può eliminare tutti gli errori che sono comuni sia al ricevitore fisso che a quello in moto, esclusi però gli errori di *multipath*, in quanto essi si verificano vicini al ricevitore, ed alcuni errori dovuti ad difetti di fabbricazione del ricevitore, in quanto essi sono propri del ricevitore stesso.
In sostanza il DGPS funziona così: noi abbiamo un ricevitore che misura gli errori cronometrici e che trasmette le informazioni corrette agli altri ricevitori in moto. In questo modo possono essere eliminati dal sistema tutti gli errori compresi quelli introdotti intenzionalmente dal Dipartimento della Difesa statunitense con la SA. L'idea è semplice, mettere il ricevitore fisso in un punto che è stato accuratamente rilevato e lasciarlo là. La stazione di riferimento riceve gli stessi segnali del ricevitore *rover*, ma invece di lavorare come un normale ricevitore GPS risolve le equazioni al contrario.
Invece di usare i segnali cronometrici per calcolare la posizione, viene usata la posizione nota per calcolare i tempi. Viene, cioè, comparato il tempo di viaggio effettivo del segnale con quello che teoricamente dovrebbe avere. La differenza tra i due tempi fornisce il fattore di correzione. Il ricevitore fisso trasmette quest'informazione d'errore al ricevitore in moto che poi la usa per correggere le informazioni.
Finché il ricevitore fisso non ha modo di conoscere quanti satelliti sono disponibili, il ricevitore in moto può essere utilizzato per calcolare la sua posizione, il ricevitore fisso procede velocemente usando tutti i satelliti visibili e computa i loro errori. Dopodiché l'informazione codificata viene trasmessa in un formato standard (ad esempio l'RTCM) ai ricevitori in moto.
I ricevitori in moto dispongono così della lista completa degli errori ed applicano le correzioni in base al particolare satellite che stanno usando.
Attualmente i ricevitori non trasmettono le correzioni autonomamente, essi sono collegati tramite opportuni ponti radio.

Le trasmissioni degli errori non includono solo gli errori di sincronia di ogni satellite, ma anche il tasso di cambiamento di tali errori. In questo modo il ricevitore in moto può interpolare la sua posizione mediante aggiornamenti. Attualmente il DGPS viene utilizzato proprio in questo modo, infatti alcune compagnie private (compagnie petrolifere o gruppi che si occupano di fare rilievi) che lavorano a grandi progetti in cui è richiesta elevata precisione, stabiliscono per i loro scopi stazioni base. Questo è diventato un approccio molto comune, viene acquistato un ricevitore fisso per poi stabilire una comunicazione con i *rover*, oggi però vi sono molte compagnie pubbliche che trasmettono le correzioni, ed è quindi possibile ottenerle gratuitamente.
Queste stazioni spesso trasmettono sui radiofari che si trovano già sul posto per le normali trasmissioni radio (normalmente intorno ai 300 KHz), chiunque in zona può ricevere queste correzioni e migliorare radicalmente la precisione delle loro misurazioni. Molte navi sono già equipaggiate di radio capaci di sintonizzarsi con i radiofari, così facendo, sfruttare il DGPS è molto facile. Molti ricevitori GPS sono stati progettati per accettare correzioni, ed alcuni sono persino dotati di ricevitori radio interni.
Non tutte le applicazioni DGPS sono uguali alcune di esse non hanno bisogno di collegamenti radio perché non hanno necessità immediata di posizionamenti precisi.
Una cosa è provare a stabilire una stazione petrolifera in un punto particolare dell'oceano, ma tutt'altra cosa è registrare il tracciato di una nuova strada per includerla in una mappa.
Per applicazioni simili alle precedenti, il ricevitore *rover* deve solo registrare tutte le posizioni misurate e il tempo esatto con cui viene effettuata ogni misurazione. Successivamente questi dati possono essere incorporati con le correzioni registrate nel

ricevitore per un ulteriore correzione dei dati. Cosi non vi è necessità di avere un collegamento radio, cosa che invece è necessaria nei sistemi in tempo reale.
Se non si ha un ricevitore di riferimento possono essere fonti alternative per ottenere le correzioni in una certa zona. Alcune istituzioni accademiche stanno sperimentando Internet come mezzo per distribuire le correzioni.
Questa è un'altra permutazione del DGPS, detta "DGPS invertito", che può essere economico per alcune applicazioni. Se per esempio noi consideriamo una linea di autobus e vogliamo individuarli sulla carta stradale con elevata accuratezza, non dobbiamo acquistare costosi ricevitori differenziali per ogni bus.
Con un sistema DGPS invertito gli autobus potranno essere equipaggiati con ricevitori GPS standard ed un trasmettitore che dovrà inviare le loro posizioni GPS standard verso l'ufficio di tracciamento. E in quest'ufficio le correzioni verranno applicate alle posizioni ricevute.
Il che richiede un computer per i calcoli ed un trasmettitore per inviare i dati, fornisce però un numero elevato di posizioni precise al costo di una stazione base, un computer ed bel po' ricevitori GPS.

4.1 DGPS con misure di codice

Tornando a quanto detto prima si ha che la stazione base calcola le correzioni di *pseudorange* PRC (*pseudorange corretion*) e le loro variazioni nel tempo RRC (*range rate corretion*). Entrambe vengono trasmesse al ricevitore remoto. La procedura può essere eseguita in tempo reale realizzando un collegamento tra le due stazioni via radio modem o modem telefonico. Il ricevitore remoto applica le correzioni alle misure *pseudorange* e calcola le posizioni punto singolo con queste osservazioni corrette, migliorando la precisione delle coordinate. Il posizionamento differenziale può essere applicato al *range* del codice o della fase.
Usando la: $R_i^j(t) = \rho_i^j(t) + c\Delta\delta_i^j(t)$, il *range* di codice sulla stazione A, misurato al epoca t_0, può essere modellato



secondo la (Hofmann-Wellenhof e al. 1997):

$$R_A^j(t_0) = \rho_A^j(t_0) + \Delta\rho_A^j(t_0) + c\delta^j(t_0) - c\delta_A(t_0)$$

dove è stato aggiunto il termine $\Delta\rho_A^j(t_0)$ errore orbitale radiale. Il *range* $\rho_A^j(t_0)$

è noto in quanto note le coordinate della stazione A (e naturalmente dei satelliti j). La correzione del *range* di codice per il satellite j all'epoca t_0 è:

$$PRC^j(t_0) = -R_A^j(t_0) + \rho_A^j(t_0) = -\Delta\rho_A^j(t_0) - c\delta^j(t_0) + c\delta_A(t_0)$$

Da una serie temporale di correzioni PRC, può essere valutata, con interpolazione numerica, la sua variazione nel tempo RRC. La correzione del *range* di codice, ad un'arbitraria epoca t, si può approssimare con la:

$$PRC^j(t_0) = PRC^j(t_0) + RRC^j(t_0)(t - t_0)$$

dove compare il termine "latenza" ($t-t_0$), determinante per la precisione di posizionamento. Questo non è altro che la differenza di tempo tra il calcolo della correzione, nella stazione base, e la sua applicazione (tempi di trasmissione calcolo ecc.) nel ricevitore remoto.

Il *range* di codice nella stazione remota B, misurato all'epoca t, può essere determinato come segue:

$$R_B^j(t) = \rho_B^j(t) + \Delta\rho_B^j(t) + c\delta^j(t) - c\delta_A(t)$$

Applicando la correzione di *range* la misura *pseudorange* in B diventa:

$$\begin{aligned} R_B^j(t)_{corretto} &= R_B^j(t) + PRC(t) = \\ &= \rho_B^j(t) + [\Delta\rho_B^j(t) - \Delta\rho_A^j(t)] + c\delta^j(t) - c\delta_B(t) - c\delta^j(t) + c\delta_A(t) = \\ &= \rho_B^j(t) + [\Delta\rho_B^j(t) - \Delta\rho_A^j(t)] - c[\delta_B(t) - \delta_A(t)] \end{aligned}$$

dove scompaiono gli errori di orologio dei satelliti. L'errore radiale d'orbita Δρ è praticamente equivalente nelle stazioni base e remota: eliminando dunque il suo contributo, si avrà:

$$R_B^j(t)_{corretto} = \rho_B^j(t) - c\Delta\delta_{AB}(t)$$

dove il termine $\Delta\delta_{AB}(t) = \delta_B(t) - \delta_A(t)$ rappresenta la combinazione degli errori d'orologio dei ricevitori.
Si noti come l'effetto perturbante della SA (degrado degli orologi e delle orbite) viene virtualmente eliminato. Stessa cosa per altri effetti di disturbo quali rifrazione ionosferica e troposferica.
La posizione del punto B è calcolata con gli *pseudoranges* di codice "corretti".
Questa correzione può essere trasmessa in formato RTCM (vedi paragrafo successivo) a partire dalla versione 2 e la tecnica è chiamata *RTCM Differential GPS*.

4.2 DGPS con misure di fase

Il *range* satellite-ricevitore misurato sulla portante, nella stazione base A all'epoca t_0 , si esprime con la (Hofmann-Wellenhof e al. 1997):

$$\lambda\phi_A^j(t_0) = \rho_A^j(t_0) + \Delta\rho_A^j(t_0) + \lambda \mathrm{N}_A^j + c\delta^j(t_0) - c\delta_A(t_0)$$

dove N_A^j è l'ambiguità di fase incognita.
Conseguentemente, la correzione del *range* della fase all'epoca t_0 sarà:

$$PRC(t_0) = -\lambda\phi_A^j(t_0) + \rho_A^j(t_0) = -\Delta\rho_A^j(t_0) - \lambda N_A^j - c\delta^j(t_0) + c\delta_A(t_0)$$

La correzione del *range* della fase all'epoca t sarà:

$$PRC(t) = PRC(t_0) + RRC(t_0)(t - t_0)$$

Seguendo la stessa procedura adottata per la correzione di codice, il range della fase della fase corretto, nel ricevitore remoto all'epoca t, si ottiene:

$$\lambda\phi_B^j(t)_{corretto} = \rho_B^j(t) + \lambda\Delta N_{AB}^j - c\delta_{AB}^j(t)$$

Il termine $\Delta N_{AB}^j = \mathrm{N}^j_B - \mathrm{N}^j_A$ rappresenta la combinazione delle ambiguità intere di fase. Anche qui vengono eliminati gli errori dell'orologio del satellite,l'effetto della SA e gli errori legati alla rifrazione ionosferica e troposferica.
La correzione del *range* della fase può essere trasmessa in tempo reale dal ricevitore base alla stazione remota tramite il protocollo RTCM a partire dalla versione 2.1, o tramite formati proprietari delle ditte costruttrici dei ricevitori.
Il DGPS con misure di fase è utilizzato per applicazioni cinematiche di precisione in tempo reale: la tecnica prende il nome RTK (Real Time Kinematic). E' richiesta in questo caso una tecnica OTF (On The Fly) per la risoluzione veloce dell'ambiguità di fase, anche durante il movimento del ricevitore.
Il posizionamento differenziale diventa equivalente a quello relativo se il tempo di latenza è nullo.

4.3 Il protocollo di trasmissione delle correzioni differenziali RTCM

La trasmissione della correzione differenziale, via radio modem o modem GSM, è stata standardizzata secondo un protocollo internazionale chiamato RTCM (*Radio Tecnical Commission for Maritime* [*service*]). La versione 2.0 contiene le sole correzioni alle misure sul *range* del codice mentre a partire dalla versione 2.1 sono presenti anche le correzioni sul *range* della fase. Il messaggio è formato da parole di 30 bit, codificate secondo record numerati (per maggiori dettagli si veda "RTCM Special Committee, 1998"). I record da 1 a 17, disponibili in tutte le versioni RTCM, contengono le correzioni di codice mentre i successivi le correzioni della fase.
I messaggi RTCM sono parte del formato internazionale NMEA, proposto da *National Marine Electronics Association for Marine Navigation*.
A titolo di esempio, si riporta l'analisi del contenuto dei massaggi di tipo 1 e 2 del protocollo RTCM, dedicati alla trasmissione della correzione differenziale sul *range* del

codice, e le strategie per la sua applicazione. La correzione differenziale di codice (*PRC*) è contenuta nel "messaggio tipo 1" del protocollo RTCM ed è associata ad ogni satellite. In generale, il *range* di codice corretto, all'epoca t (*PR(t)*), si ottiene correggendo lo *pseudorange* misurato (*PRM(t)*):

$$PR(t) = PRM(t) + PRC(t) = PRM(t) + PRC(t_o) + RRC(t- t_o)$$

Nello stesso messaggio viene fornita una stima dell'accuratezza (1σ) della correzione *pseudorange*, detta UDRE (*User Differential Range Error*) che contiene gli errori combinati di *multipath*, *noise* e altri effetti. Una sezione del messaggio 1 è rappresentata nella tabella seguente.

1	UDRE	Sat.ID	PRC	RRC	IOD	1	UDRE
bit	2 bit	5 bit	16 bit	8 bit	8 bit	bit	2 bit

Il messaggio 1 contiene dunque tutte le correzioni per ogni satellite visto dalla stazione base: sono necessari fino a 40 bit per ogni satellite.
Se durante la misura cambia la costellazione, il valore del segmento IOD (*Issue Of Data*) permette di valutare se il calcolo di posizione nella stazione remota è le correzioni nella stazione base, derivano dalla stessa configurazione satellitare. Fin quando il ricevitore remoto non decodifica dalle effemeridi la nuova costellazione, ne utilizza una diversa da quella con cui la stazione base calcola la correzione differenziale: questo può causare errori di posizionamento. In questo caso viene utilizzato il messaggio RTCM tipo 2 che contiene la differenza sulla correzione *pseudorange* (*ΔPRC*) e sul *range rate correction* (*ΔRRC*) causate dal cambiamento della costellazione. La stazione base esegue due calcoli di PRC ed RRC con costellazione vecchia (*old IOD*) e nuova (*new IOD*); le differenze:

$$\Delta PRC = PRC(old\ IOD) – PRC(new\ IOD)$$

$$\Delta RRC = RRC(old\ IOD) – RRC(new\ IOD)$$

Vengono applicate nel ricevitore remoto per ricavare la correzione differenziale al tempo t:

$$PRC(t) = PRC(new\ IOD) + \Delta PRC(old\ IOD) + RRC(new\ IOD)(t-t_1)+ \Delta RRC(old\ IOD)(t-t_2)$$

Dove $(t-t_2)$ rappresenta la latenza del messaggio tipo 2.
La struttura del messaggio tipo 2 è del tutto analoga a quella del messaggio tipo 1.
La correzione differenziale può essere ricavata a posteriori o ricevuta via radio e/o GSM da una stazione base posta nella zona di intervento oppure da stazioni permanenti GPS (se questo servizio è attivato).
Attualmente può essere ricevuta anche da satellite, per esempio quelli della rete OMNISTAR, e applicata in tempo reale alle misure di codice.

4.4 L'RTK

4.4.1 Generalità sull'RTK

Il DGPS con misure di fase in tempo reale prende il nome di RTK. Le misure di fase funzionano come un nastro metrico con gradazione millimetrica. Utilizzando le misure di fase non compaiono nell'immediato i valori metrici, bisogna attendere per un certo lasso temporale (un po' come avviene per le foto POLAROID), durante questo intervallo vengono risolte le ambiguità di fase. Maggiore è l'intervallo temporale di attesa più precisa sarà la risoluzione dell'ambiguità. Dopo che i valori metrici sono stati restituiti è possibile fare misurazioni immediate per tutto il tempo del rilievo. Se si verificano *cycle slip* vengono persi i valori metrici e bisogna ripetere la risoluzione dell'ambiguità.
Nel caso in cui la perdita d'agganciamento con i satelliti duri solo pochi istanti, il ricevitore può fissare il *cycle slip* basandosi sul calcolo dell'ambiguità della precedente epoca.
Se però l'attribuzione degli interi non è stata eseguita correttamente o si presentano mancanze di cicli è come aver letto erroneamente le gradazioni dal nastro metrico, ad esempio è come aver calcolato una distanza di 4,784 metri mentre essa è invece pari a 3,784. La gradazione millimetrica (numero frazionario di lunghezze d'onda o di cicli) è stata letta bene quella metrica (numero intero di lunghezze d'onda o di cicli) invece no.
Quando un ricevitore risolve l'ambiguità di fase la precisione nel calcolo della posizione è compresa tra 0,5 e 2 cm orizzontalmente e tra 1 e 5 cm verticalmente (tutto dipende dalla capacità dell'antenna di eliminare errori *multipath*) considerando un'incertezza di 1 ppm per ricevitori a doppia frequenza e di 2 ppm per ricevitori a singola frequenza.
Come per tutte le altre tecniche di rilievo GPS la chiave dei rilievi RTK è la risoluzione dell'ambiguità, la domanda più importante che bisogna porsi è: "quanto tempo bisogna impiegare per ottenere una risoluzione affidabile?".
Per *baseline* più corti di 10 km il tempo richiesto per risolvere l'ambiguità dipende da alcuni parametri:

- Livello di precisione scelto per il rilievo e conseguentemente per la risoluzione del numero dei cicli interi.
- Numero dei satelliti osservati (o disponibili per l'osservazione).
- Caratteristiche dei ricevitori (singola o doppia frequenza).
- Contributo di errore del *multipath* (coefficiente di riflessione del terreno).
- Capacità dell'antenna del ricevitore di eliminare il *multipath*.

Il tempo richiesto per la risoluzione dell'ambiguità viene nel seguito descritto considerando singolarmente i parametri per poterne valutare individualmente l'importanza.

Se tutti i parametri hanno la stessa importanza è probabile che servano 10 s per risolvere l'ambiguità con una affidabilità del 99%, e 100 s con un'affidabilità del 99,9%.

Se tutti gli altri parametri hanno la stessa importanza è probabile che basti 1 s per risolvere l'ambiguità osservando 15 satelliti, il tempo sale a 100 s con 8 satelliti e a 10 min. con 6 satelliti.

Se tutti gli altri parametri hanno la stessa importanza è probabile che servano 10 s utilizzando ricevitori a doppia frequenza e 2 min. utilizzando ricevitori a singola frequenza. Utilizzando ricevitori a doppia frequenza è come se si disponesse del 50% di satelliti in più rispetto alla situazione analoga con ricevitori a singola frequenza. Ad esempio è come se

io, utilizzando ricevitori a doppia frequenza (moltiplicatore 1,5 per il numero di satelliti), osservassi 9 satelliti pur disponendo effettivamente di 6 satelliti.

Se tutti gli altri parametri hanno la stessa importanza è probabile che servano 10 s per misurazioni a terreno asciutto e 20 s a terreno bagnato. A terreno bagnato le riflessioni sono molto più intense.

Se tutti gli altri parametri hanno la stessa importanza è probabile che servano 10 s per risolvere l'ambiguità, disponendo di un'antenna che riduca abbastanza l'effetto del *mutipath* (-35 db down/up ratio) e 200 s con un'antenna normale (-15 db down/up ratio). Il *multipath* la causa principale della erronea valutazione del numero di cicli interi.

Considerando lunghi *baseline* sono richiesti almeno 8 satelliti e ricevitori a doppia frequenza.

Se consideriamo il solo sistema GPS 6 satelliti sono visibili per il 96% della giornata, (questo significa un numero soddisfacente di satelliti per rilievi con *baseline* corti) 8 satelliti sono visibili per il 32% della giornata, e solo il 5% della giornata per 9 satelliti.

Se oltre al GPS consideriamo anche i satelliti della costellazione GLONASS[13] 9 satelliti sono osservabili per il 99% della giornata, e più di 11 per 10,96% della giornata.

La possibilità che vengano eseguiti rilievi RTK con ricevitori a singola frequenza ed utilizzando solo satelliti GPS (5% della giornata) può essere rara se consideriamo *baseline* corti e non si verifica mai se consideriamo *baseline* lunghi.

La possibilità che vengano eseguiti rilievi RTK con ricevitori a singola frequenza sfruttando satelliti GPS e GLONASS (99,5% della giornata) è considerata soddisfacente per piccole basi, tale possibilità non viene presa in considerazione per grandi basi.

La possibilità che vengano eseguiti rilievi RTK con ricevitori a doppia frequenza sfruttando i soli satelliti GPS (96% della giornata) è considerata soddisfacente per piccole basi, tale possibilità non viene presa in considerazione per grandi basi.

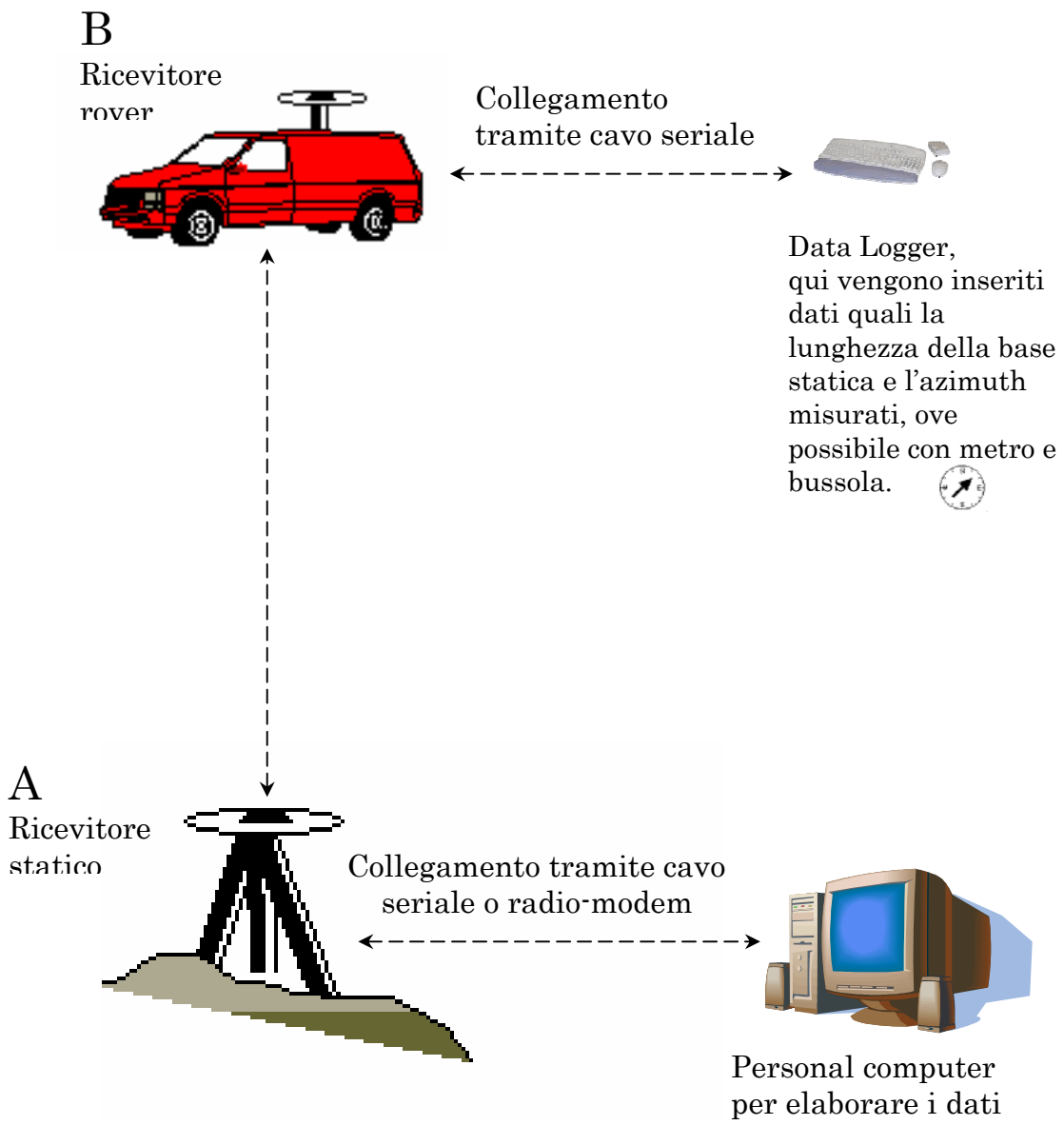
Combinando osservazioni GPS e GLONASS ed utilizzando ricevitori a doppia frequenza, si può operare sempre ed in modo soddisfacente per i rilievi RTK, sia per piccole che per grandi basi.

4.4.2 Strumentazione utilizzata per l'RTK

Il metodo RTK (*Real Time Kinematic*) consiste nella trasmissione in tempo reale via cavo seriale, oppure, a distanza, via radio modem dei dati grezzi *pseudorange* e di fase, oltre che delle coordinate impostate dall'operatore, da un ricevitore geodetico fisso che funge da *master* statico (*static*) ad un altro ricevitore geodetico che è previsto in movimento (*rover*), dopo una breve sosta di inizializzazione. Quest'ultimo esegue: il calcolo delle ambiguità iniziali elaborando i dati della base statica iniziale, che possono essere forniti o direttamente dalla tastiera frontale dello strumento, impostando come coordinate di

[13] Il GLONASS è la controparte ex-sovietica del sistema satellitare GPS, una delle differenze più importanti tra le due costellazioni consiste nel fatto che mentre i satelliti GPS hanno tutti la stessa frequenza fondamentale (f_0=10,23 MHz), nel GLONASS la frequenza fondamentale varia da satellite a satellite.

riferimento quelle calcolate grazie ad una preventiva osservazione statica della *baseline master-rover*, oppure attraverso un *data-logger* collegato tramite un cavo seriale allo strumento *rover*; in questo caso si può impostare la lunghezza e l'azimuth della base statica, misurati con metro e bussola se la base è molto corta (inferiore a 2 metri) e orizzontale; alternativamente è possibile impostare le coordinate delle due antenne, ma allora è più conveniente operare direttamente sulla tastiera dei due strumenti dato che cosi di solito si ha in disposizione un'uscita seriale in più libera che può essere utilizzata per l'output dei vari messaggi (*NMEA, Cycle Printout*) dal ricevitore mobile, per la memorizzazione su PC, e l'eventuale visualizzazione su schermo. La registrazione dei dati dei due ricevitori è comunque possibile anche nella loro memoria interna o attraverso il *data-logger*. Quest'ultimo consente inoltre di operare direttamente le trasformazioni di coordinate, utili per le applicazioni cartografiche.



Il ricevitore mobile, quindi, esegue la risoluzione del sistema di equazioni di osservazione alle differenze seconde, in tempo reale, epoca per epoca; ciò rende possibile il rilievo di traiettorie oltre che il rilievo di coordinate di punti con il metodo Stop & Go. Il ritardo di

presentazione dei risultati nelle varie uscite è di circa 4 secondi, dovuti ai tempi di elaborazione.

Questa strumentazione consente di effettuare vari tipi di operazioni fra le quali:

- elaborazione in tempo reale e registrazione delle coordinate di vertici con il metodo GPS cinematico Stop & Go;
- elaborazione in differita di tutto il percorso effettuato con il metodo Stop & Go, ciò anche dei tratti fra i vertici, per la delimitazione di aree e altre applicazioni;
- elaborazione di una base statica e visualizzazione immediata della distanza azimuth, dislivello fra i due caposaldi e delle coordinate del punto estremo della baseline;
- tracciamento centimetrico, anche per batimetrie, con la procedura di navigazione disponibile sia sul display dello strumento che sul quello del *data-logger*;
- registrazione sul PC delle coordinate del *Rover* fermo o in movimento, ogni secondo o con intervallo a piacere, anche remotamente, via radio- modem, in formato compatto con risoluzione millimetrica. Questa modalità può essere utilizzata sia nelle misure di controllo in tempo reale di deformazioni del suolo o di strutture, sia nel tracciamento di veicoli.

In più e comunque sul display del ricevitore *Rover* sono visualizzate, epoca per epoca, le componenti cartografiche E, N ed il dislivello ellissoidico (o geoidico) della *baseline master-rover.*
La registrazione della variazione delle componenti della base fra i due ricevitori e la sua visualizzazione in tempo reale sullo schermo di un calcolatore portatile, consente di fare delle valutazioni, in sito, dell'entità del rumore dovuto ad effetti locali (in particolare effetti atmosferici e *multipath*), il che consente di calcolare le precisioni ottenibili è l'eventualità di modificare la localizzazione del *master*.
È anche possibile individuare forti irregolarità dovute a particolarità geometriche (valori elevati dell'RDOP) che consiglino di variare l'azimuth della *baseline master-rover*.
La possibilità di ottenere contemporaneamente in uscita dal *master* le correzioni *pseudorange* RTCM consente di operare in DGPS *pseudorange* con altri ricevitori.

4.4.3 Tecniche di rilievo RTK/ OTF

Il rilievo RTK/OTF è simile al rilievo GPS cinematico differenziale in quanto sono richiesti due ricevitori che registrano simultaneamente le osservazioni, è ammessa la possibilità che il ricevitore *rover* effettui il rilievo in movimento. A differenza del rilievo cinematico, le tecniche di rilievo RTK/OTF usano osservazioni GPS a doppia frequenza L1/L2 ed è tollerata la possibilità che si perda il segnale dal satellite. Dato che l' RTK/OTF usa la frequenza L2, il ricevitore GPS deve essere in grado di tracciare la frequenza L2 durante l'A-S. Ci sono varie tecniche usate per ottenere L2 durante l'Anti-Spoofing. Esse includono metodi di quadratura e di correlazione incrociata.

a) Risoluzione dell'ambiguità. La risoluzione soddisfacente dell'ambiguità è richiesta per la formulazione soddisfacente dei *baseline*. La tecnologia RTK/OTF consente al ricevitore remoto di inizializzare e risolvere questi interi senza un periodo di inizializzazione statica. Con l'RTK/OTF, se si verificano *cycle slip* l'inizializzazione può avvenire in moto. Gli interi possono essere risolti dal *rover* entro 10 ÷ 30 secondi, tutto dipende dalla distanza dalla stazione di riferimento. L' RTK/OTF usa la frequenza L2 trasmessa tramite i satelliti GPS

nella risoluzione dell'ambiguità. Dopo che gli interi sono stati risolti, solo L1 e C/A sono utilizzati per computare le posizioni.

b) Procedura di rilievo. Il rilievo RTK/OTF richiede ricevitori GPS a doppia frequenza L1/L2. Uno dei ricevitori viene posto su un punto noto, e l'altro viene installato su una piattaforma mobile o in moto. Se il rilievo viene effettuato in tempo reale, un collegamento per trasferire i dati ed un processore (esterno o interno) sono richiesti. Il collegamento è usato per trasferire i dati grezzi dalla stazione di riferimento alla stazione *rover.*

- *Processore interno.* Se il sistema RTK/OTF viene eseguito con un processore interno (presente nel ricevitore), si devono seguire le linee guida del costruttore.
- *Processore esterno*. Se l'RTK/OTF è effettuato con processori esterni (ad esempio computer portatili), allora il computer nella stazione di riferimento (un PC basato sul processore 386) raccoglie i dati grezzi GPS, li formatta e li invia tramite il collegamento dati al ricevitore remoto. Il computer al ricevitore *rover* (un PC basato sul processore 486 a 33 MHz) processa i dati grezzi provenienti dalla stazione di riferimento e da quella remota per risolvere gli interi ed ottenere una posizione.

c) Precisione dei rilievi RTK/OTF. I rilievi RTK/OTF sono accurati entro i 10 cm quanto la distanza fra *rover* e *master* non supera i 20 km. I risultati dei test hanno riscontrato precisioni inferiori al livello decimetrico.

4.5 Elaborazioni in doppia frequenza

Le elaborazioni in doppia frequenza possono essere eseguite in svariate modalità, combinando opportunamente le frequenze L1 e L2. Considerando alcune di queste combinazioni:

Combinazione "WIDE-LANE"

E' una combinazione che viene normalmente impiegata per il fissaggio dell'ambiguità di fase. Si è visto come nelle doppie differenze si ricerca l'intero più vicino: a causa del rumore di misura la ricerca può dare risultati diversi nel tempo. Se il rumore di misura è dell'ordine di grandezza della lunghezza d'onda, si può intuire la difficoltà ad individuare un solo probabile intero.
Conviene allora utilizzare combinazioni delle portanti che permettano di simulare lunghezze d'onda maggiori. Si consideri la combinazione *wide-lane* la cui frequenza L_W è ottenuta come differenza delle frequenze L1-L2:

$$L_W = L1 - L2 = 1575.42MHz - 1227.60MHz = 347.82MHz$$

La lunghezza d'onda risultante sarà: $\lambda_W = \frac{C}{L_W} = \frac{3 \times 10^8 m/s}{347.82 \times 10^8 Hz} \cong 86cm$

E' più facile determinare l'ambiguità di fase se la lunghezza d'onda è maggiore, specie in presenza di rumori di misura, *multipath* o altro.

Combinazione "NARROW-LANE"

Ha una lunghezza d'onda di circa 11 cm: è più precisa ma è più difficile determinare l'ambiguità intera di fase. Deriva dalla combinazione:

$$L_N = L1 + L2 = 1575.42MHz + 1227.60MHz = 2803.02MHz$$

La lunghezza d'onda risultante sarà: $\lambda_N = \frac{C}{L_N} = \frac{3 \times 10^8 m/s}{2803.02 \times 10^8 Hz} \cong 11cm$

Combinazione "IONO-FREE"

La ionosfera contiene elettroni liberi provenienti da molecole ionizzate dalle radiazioni solari. La densità elettronica aumenta con la quota fino a 300-400 km e successivamente decresce; questo andamento dipende dall'intensità della radiazione solare che decresce al diminuire della quota e dalla densità atmosferica che, viceversa, decresce all'aumentare della quota. Numerosi sono i fenomeni che influenzano il numero di elettroni liberi; attività solare, posizione geografica e geomagnetica, stagione e momento durante la giornata.
La propagazione di un segnale elettromagnetico che attraversa la ionosfera viene influenzata in due modi differenti: la sua velocità di fase aumenta mentre la sua velocità di gruppo diminuisce; per questo motivo, le misure di distanza derivate da misura di fase per effetto della, ionosfera risultano "accorciate" mentre quelle derivate da misure di codice (basate sulla velocità di gruppo) risultano "allungate". I due effetti hanno uguale entità (a. parità di frequenza) ma segno opposto; trascurando termini superiori al secondo ordine, l'indice di rifrazione del segnale è esprimibile come:

$$n = 1 \mp \frac{AN_e}{f^2}$$

dove:

A è un valore costante che vale 40.3 m /s;
f è la frequenza del segnale;
N è la densità di elettroni;

si noti che la variazione dell'indice di rifrazione è tanto più elevata quanto minore è la frequenza del segnale. La variazione di distanza, in valore assoluto, risulta pertanto:

$$\delta\rho_{ion} = \int_{ion} (n-1) d\rho = \frac{A}{f^2} \int_{ion} N_e d\rho$$

l'integrale della densità elettronica esteso a tutto il percorso interno alla ionosfera è detto contenuto totale di elettroni (TEC) e, calcolato relativamente ad una visuale zenitale, vale in genere 10^{16}-10^{18} el/m^2:

$$\delta\rho_{ion} = \frac{A(TEC)}{f^2} = \frac{a}{f^2}$$

In particolare esso può variare anche di un ordine di grandezza tra la notte ed il giorno, cosicché relativamente alle frequenze delle portanti impiegate nel sistema GPS le variazioni giornaliere di $\delta\rho_{ion}$ raggiungono facilmente i 10 m (mediamente l'effetto ionosferico è 4 volte maggiore durante il giorno rispetto alla notte). In casi estremi, per un valore massimo di (TEC) di $2*10^{18}$ el/m^2, allo zenith $\delta\rho_{ion}$ raggiunge i valori di 32 m e 53 m rispettivamente per le frequenze LI e L2, e per angoli di elevazione inferiori aumenta ulteriormente; la corrispondente variazione del tempo di percorso è dell'ordine di 100-150 ns. L'errore che si commette trascurando i termini di ordine superiore al secondo in tali condizioni, vale rispettivamente 3.6 cm e 7.6 cm per le due portanti LI e L2.
Se consideriamo ora due ricevitori che osservano il medesimo satellite essi, in generale ricevono i segnali con angoli di elevazione diversi, e quindi l'effetto della ionosfera risulta diverso; in particolare le distanze tra il satellite ed i due ricevitori vengono "accorciate" di quantità differenti e ciò introduce un effetto di scala sulla lunghezza della base definita dai due ricevitori che è dell'ordine di $0.5*10^{-6}$ per un (TEC) di 10^{17} el/m^2, e quindi non è affatto trascurabile.
Vi sono diverse strategie per trattare il *bias* ionosferico; la scelta dipende dai tipi di ricevitori a disposizione (singola o doppia frequenza) e quindi, in ultima analisi, dalle precisioni che si vogliono ottenere.
In primo luogo, il *bias* ionosferico può essere eliminato, almeno fino al secondo ordine, combinando opportunamente osservazioni compiute su segnali aventi frequenze diverse se tutti i ricevitori utilizzati sono a doppia frequenza; si noti che in questo caso per frequenza si intende sempre quella dell'onda portante (L1 o L2), anche se le osservazioni si eseguono relativamente al codice P; il codice C/A è inutilizzabile in questa operazione in quanto modula esclusivamente la portante L1 e, del resto, la sua mediocre precisione non giustifica l'adozione della correzione del *bias* ionosferico). Si considerino per semplicità misure di *pseudorange* e di fase affette dal solo *bias* ionosferico; tale ipotesi non toglie generalità al discorso poiché gli altri effetti che influenzano le misure sono indipendenti da questo *bias* e vengono modellizzati a parte sia nelle misure originarie che in quelle ottenute da una loro combinazione.
Per quanto guarda le misure di *pseudorange* con codice P, sulla base dell'ipotesi precedente, si può scrivere:

$$\begin{cases} P^s_{r\,L1} = \rho^s_r + \dfrac{a}{f^2_{L1}} \\ P^s_{r\,L2} = \rho^s_r + \dfrac{a}{f^2_{L1}} \end{cases} \Rightarrow \begin{cases} P^s_{r\,L1} = \cdots\cdots\cdots\cdots \\ P^s_{r\,L2} - P^s_{r\,L1} = a\left(\dfrac{1}{f^2_{L2}} - \dfrac{1}{f^2_{L1}}\right) \end{cases} \Rightarrow \begin{cases} P^s_{r\,L1} - \dfrac{\left(P^s_{r\,L2} - P^s_{r\,L1}\right) f^2_{L2}}{f^2_{L1} - f^2_{L2}} = \rho^s_r \\ a = \dfrac{\left(P^s_{r\,L2} - P^s_{r\,L1}\right) f^2_{L2}\, f^2_{L1}}{f^2_{L1} - f^2_{L2}} \end{cases} \Rightarrow \begin{cases} \dfrac{P^s_{r\,L1} f^2_{L1} - P^s_{r\,L2} f^2_{L2}}{f^2_{L1} - f^2_{L2}} = \rho^s_r \\ a = \cdots\cdots\cdots\cdots \end{cases}$$

dopo aver risolto il sistema di cui sopra nelle incognite a e ρ^s_r si ottiene la combinazione che permette di determinare la distanza satellite-ricovitore e l'effetto ionosferico:

$$P^s_{r\;iono-free} = \alpha_P\, P^s_{r\,L1} + \beta_P\, P^s_{r\,L2} = \rho^s_r$$

dove:

$$\alpha_P = \frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2}$$

$$\beta_P = -\frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2}$$

Per le misure di fase, se si tiene presente che il termine ionosferico si può scrivere come:

$$\frac{a}{f^2\lambda} = \frac{1}{c}\frac{a}{f}$$

da cui è possibile ricavare una posizione che utilizzeremo più avanti: $\lambda = \frac{c}{f}$

l'equazione di osservazione diventa allora:

$$\lambda\Phi_r^s = \rho_r^s - \lambda N + \frac{a}{f^2}$$

e quindi per le due frequenze si ha:

$$\begin{cases} \Phi_{rL1}^s = \frac{f_{L1}}{c}\rho_r^s - N_{L1} + \frac{a}{cf_{L1}} \\ \Phi_{rL2}^s = \frac{f_{L2}}{c}\rho_r^s - N_{L2} + \frac{a}{cf_{L2}} \end{cases}$$

In modo del tutto analogo a quanto fatto per le misure di *pseudorange*, la soluzione del precedente sistema rispetto ai parametri a e ρ_r^s (quest'ultimo a meno di ima combinazione delle ambiguità di fase) fornisce la combinazione lineare indipendente dall'effetto ionosferico:

$$\frac{\rho_r^s}{c} - \left[\frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2}N_{L1} - \frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2}N_{L2}\right] = \frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2}\Phi_{r_{L1}}^s - \frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2}\Phi_{r_{L2}}^s$$

per ottenere dei coefficienti adimensionali si possono moltiplicare entrambi i mèmbri per f_{L1} e quindi l'equazione diventa:

$$\Phi_{r\,iono-free}^s = \alpha_\Phi \underset{L1}{\Phi_r^s} + \beta_\Phi \underset{L2}{\Phi_r^s} = \frac{f_{L1}^2}{c}\rho_r^s - \left[\alpha N_{L1} + \beta N\right]_{L2}$$

L'ambiguità iniziale di fase di questa equazione non è più una quantità intera, in quanto combinazione lineare a coefficienti non interi delle ambiguità relative alle due portanti N_{L1}, N_{L2}; ciò costituisce un ostacolo alla ricerca dei valori interi di tali ambiguità, poiché coppie diverse possono generare valori molto simili, e quindi difficilmente distinguibili, del l'ambiguità complessiva che viene stimata; un problema del tutto analogo si presenta nella rimozione dei *cycle slip*. Per superare questa difficoltà, sono stato proposte due

combinazioni lineari a coefficienti interi delle misuro di fase per eliminare il *bias* ionosferico:

$$\Phi^{s}_{r43} = 4\Phi^{s}_{rL1} - 3\Phi^{s}_{rL2}$$

$$\Phi^{s}_{r54} = 5\Phi^{s}_{rL1} - 4\Phi^{s}_{rL2}$$

Utilizzo delle combinazioni di fase

La soluzione *wide-lane* non è in generale quella ottimale ma serve a stimare le ambiguità intere. Determinati questi valori di ambiguità, si procede con altre soluzioni tipo *iono-free*, *narrow-lane* o solo L1. Si analizza la validità delle soluzioni:

- *Soluzione iono-free FIX* (con ambiguità fissata ad intero): è quella ottimale nella maggior parte dei casi. Ottima specialmente quando si stimano le ambiguità intere con *wide-lane;*
- *Soluzione iono-free FLT:* non è generalmente quella ottimale per basi corte (< 30 Km). Va bene per le basi di centinaia di km con lunghi tempi di acquisizione (su basi molto lunghe si arriva poi al fissaggio delle ambiguità);
- *Soluzione wide-lane FIX:* è una soluzione da utilizzare con cautela (può andar bene per basi di 15 ÷ 30 km). Rappresenta normalmente un passaggio intermedio per il calcolo dell'ambiguità e il successivo calcolo *iono-free*. (non è generalmente consigliabile la soluzione *wide-lane FLT*)
- *Soluzione L1 FIX:* da usare su basi corte (minori di 15 km ma più spesso 5 ÷ 10 km) dove la ionosfera non influisce fortemente. La soluzione finale può essere preceduta dalla soluzione delle ambiguità con *wide-lane*;
- *Soluzione L1 FLT:* sopra 10 ÷ 15 km di lunghezza della base, è difficile arrivare a fissare gli interi in singola frequenza: è necessaria la doppia frequenza. Le soluzioni L1 FLT sono meno precise di quelle L1 FIX. Per distanze maggiori di 15 km questa soluzione è poco affidabile a causa dell'effetto del ritardo ionosferico. Un altro motivo per cui si arriva a questa soluzione è un tempo di acquisizione troppo breve per il fissaggio dell'ambiguità di fase. Sono in generale sconsigliabili.

4.6 Elaborazioni in singola frequenza

La soluzione ottimale è quella con ambiguità fissate ad intero (FIX). Se la soluzione non passa il test del ratio, viene generata la soluzione ad interi non risolti (FLT). La soluzione FLT sulla sola L1 non è una soluzione ottimale: spesso il motivo per cui non si raggiunge il fissaggio dell'ambiguità è un tempo di occupazione troppo breve.
Per basi superiori a 15 km, diventa difficile arrivare alla soluzione FIX in singola frequenza, a causa della sensibilità delle osservazioni ai ritardi ionosferici, non eliminabili in singola frequenza.

Esempio di soluzione di base GPS - singola base.

From Station	**To Station**	**Solution**	**Slope**	**Ratio**
3	**903**	**Iono free fixed**	**23087.240**	**19.0**
Occupation Time	Meas. Interval (seconds):	**03.54.00,00**		**15.00**
Solution Type:		Iono free fixed double difference		
Solution Accettability:		Passed ratio test		
Ephemeris:		Broadcast		
Baseline Slope Distance (meters):		**23087.240**		**0.000444**
Baseline Components (meters):		**dx -10446.327**	**dy 19061.009**	**dz 7782.885**
Standard Deviations (meters):		**0.000839**	**0.000408**	**0.000918**
		dn 10981.643	de 20308.205	du -31.431
		0.000624	0.000404	0.001077
				dh 10.314
				0.001077
A posteriori Covariance Matrix:		**7.038408E-007**		
		4.756531E-008	**1.662718E-007**	
		3.904630E-007	**-2.534002E-009**	**8.433269E-007**

Dall'analisi di questa soluzione, risulta che la base tra i vertici 3 e 903, di circa 23 km, è stata risolta con la soluzione ad ambiguità fissata ad intero, nella combinazione libera dall'effetto ionosferico (*Iono free Fixed*). Il fissaggio dell'ambiguità di fase è avvenuto con succeso essendo il ratio un valore elevato pari a 19. Viene riportato lo scarto quadratico medio (*Standard Deviations*) del vettore di base (*Slope Distance*) le componenti della base (dx, dy, dz) del sistema geocentrico WGS84 e nel datum locale (de, dn, du) con i relativi scarti quadratici medi ed infine la matrice di varianza covarianza della base.

4.7 Il WAAS - un'applicazione del DGPS

La FAA (*Flight American Aviation*) ha ottenuto grandi benefici per l'aviazione usando il GPS, la precisione del DGPS è a disposizione degli aerei che viaggiano in tutto il continente americano. Il loro progetto è noto con il nome di WAAS (*Wide Area Augmentation System*, trad. Sistema d'incremento per zone estese), ed è fondamentalmente un sistema DGPS continentale.
L'idea è nata dal fatto che alcune richieste specifiche non potevano essere soddisfatte con l'impiego del solo GPS, tutto è iniziato dal "problema dell'integrità di sistema". Il GPS è molto affidabile, ma ogni tanto il malfunzionamento di un satellite causa l'invio dati poco precisi. Le stazioni di monitoraggio individuano questo genere di problemi ed inviano un messaggio sullo stato del sistema in cui viene "detto" ai ricevitori di ignorare il satellite rotto fino a nuovo ordine. Sfortunatamente questo processo può durare alcuni minuti, per un aereo in atterraggio questo è un intervallo temporale troppo grande.
La FAA ha allora pensato di mettere in piedi un proprio sistema di monitoraggio che reagisse molto più velocemente, hanno infatti pensato di mandare in orbita un satellite geosincrono sopra gli Stati Uniti che può allertare istantaneamente gli aerei quando si verifica un problema di questo tipo. L'informazione viene quindi trasmessa sul canale GPS in modo che gli aerei sono in grado di riceverla sui normali ricevitori senza aver bisogno di apparati radio aggiuntivi.

Pensarono poi ad utilizzare questo satellite anche per scopi di posizionamento e, dopo aver inviato in orbita un altro satellite geosincrono (allo scopo di infittire la costellazione del loro sistema) e di utilizzarli entrambi per trasmettere correzioni differenziali.
La FAA ha poi installato una rete di 24 ricevitori di riferimento fissi in tutti gli Stati Uniti, disponendo così di dati corretti per la maggior parte del territorio nazionale. Questi dati rendono il GPS sufficientemente accurato per atterraggi di classe 1 (cioè molto vicini alla pista d'atterraggio, ma non con visibilità zero).
Le implicazioni di questo sistema vanno ben oltre l'aviazione dato che le correzioni saranno disponibili per tutti gli utenti. Inoltre, per completare il sistema, la FAA intende stabilire un LAAS (*Local Area Augmentation System* trad. Sistema d'incremento per zone locali) vicino alle piste, esso funziona come il WAAS ma su scala ridotta, i ricevitori vicino alla pista forniranno le posizioni agli aerei in atterraggio.
Con il LAAS il GPS sarà impiegabile per effettuare atterraggi di classe 3 (cioè con visibilità zero).

4.8 Gli algoritmi di risoluzione OTF[14]

4.8.1 Risoluzione dell'ambiguità "On the Fly"

La base delle moderne tecniche di posizionamento GPS consiste nella capacità di convertire il trattamento dei dati da un problema che verte sulle misure di fase osservate ad uno che risolve le più semplici misure di *pseudorange*.
Questo risultato viene raggiunto tramite un processo matematico di risoluzione dell'ambiguità, che consiste nella determinazione dei valori interi dei parametri ambigui delle doppie differenze. La risoluzione dell'ambiguità è quindi un processo con cui le misure di fase osservate, precise ma ambigue, vengono convertite in quantità *range* non ambigue. Tali quantità hanno tutti i vantaggi delle osservazioni *pseudorange* ma hanno rumore di misura molto più basso.
Varie tecniche sono state progressivamente raffinate al fine di raggiungere questi risultati, tant'è che oggi è possibile applicarle mentre il ricevitore è in moto (questa è la cosiddetta "modalità On the Fly"), entro periodi estremamente corti, pochi secondi o minuti.
La procedura di risoluzione dell'ambiguità consiste in vari stadi:

(1) Definizione a priori del valore dei parametri di ambiguità.
(2) Azione di un algoritmo di ricerca per identificare i valori interi più probabili.
(3) Implementazione di un algoritmo decisionale per scegliere il "migliore" set di valori interi.
(4) Applicazione delle ambiguità risolte ai dati per ottenere *range* misurati non ambigui.

Idealmente i primi tre passi possono avvenire in modo trasparente per l'utente, con poco ritardo ed elevata affidabilità. Varie tecniche sono state sviluppate per attuare i passi (2) e (3), alcune di esse sono:

- La FARA (*Fast Ambiguity Resolution Approach* - Approccio per la Risoluzione Veloce dell'Ambiguità).
- La tecnica di ricerca basata sulla Decomposizione Cholesky.
- La tecnica di ricerca basata sulla Decomposizione Spettrale.
- La tecnica di ricerca delle ambiguità ai minimi quadrati.

[14] Per una trattazione maggiormente dettagliata degli algoritmi OTF convenzionali si vada all'Appendice E

- Il FASF (*Fast Ambiguity Search Filter* – Filtro di Ricerca Veloce dell'Ambiguità).
- La tecnica LAMBDA (*Least Squares AMBiguity Decorrelation Adjustement* - Arrotondamento ai Minimi Quadrati della decorrelazione dell'Ambiguità).
- Il Metodo-Funzione Ambiguità.
- Stima diretta da una combinazione di misure di codice e di fase.

Nessuna delle procedure sopra elencate è al 100% a prova d'errore, quindi è necessario trovarsi in condizioni operative favorevoli per ottenere risultati corretti:

- Fare in modo che la distanza tra i due ricevitori sia abbastanza corta. *Generalmente meno di 20 km*.
- Tracciare quanti più satelliti possibile per assicurare una buona geometria satellitare. *Preferibilmente cinque o più satelliti*.
- Utilizzare osservazioni a doppia frequenza. *In modo che sia possibile usare valide combinazioni di dati*.
- Utilizzare dati *pseudorange* precisi, *in quanto aiutano ad ottenere a priori valide informazioni di posizionamento*.
- Migliorare la ricerca dell'ambiguità, scegliendo e testando gli algoritmi. *E' su questo punto che si concentra molta ricerca*.

Il "Santo Graal" del posizionamento preciso consiste nell'ottenere la risoluzione istantanea dell'ambiguità (cioè in una singola epoca). Sebbene siano stati fatti progressi significativi in questo senso, non si dispone ancora di un processo di routine affidabile.
Per applicare le tecniche note nel miglior modo possibile è quindi necessario disporre di hardware GPS che rappresenti lo "stato dell'arte" in quanto a tecnologia, tranne che queste tecnologie non si possono considerare disponibili per la maggior parte delle applicazioni GPS standard.

4.8.2 Algoritmi basati sul filtro di Kalman[15]

L'uso del filtro di Kalman per il trattamento dei dati GPS è una tendenza in crescita, ma è solo una trovata di mercato?
Il metodo standard dei minimi quadrati è tipicamente usato quando il (sistema) problema di stima è sovradeterminato o, in altre parole, quando sono disponibili più osservazioni di quante in effetti ne servano per stimare i parametri di posizione. Nelle applicazioni cinematiche il metodo dei minimi quadrati può essere applicato ai dati su una base "epoca per epoca".
I parametri d'interesse (la posizione), e gli errori di sistema dominanti (per esempio, gli errori d'orologio e quelli di rifrazione atmosferica), sono quantità variabili rispetto al tempo. Per tali applicazioni, le tecniche di trattamento dei dati che sono più efficienti ed ottimali, e quindi le più appropriate, sono quelle basate sull'estensione dei principi dei minimi quadrati per racchiudere i concetti di *previsione*, *filtraggio* ed *arrotondamento*. Questi ultimi tre concetti sono strettamente correlati e possono essere illustrati tramite un esempio:

Si consideri un veicolo in moto i cui parametri d'interesse siano le sue posizioni istantanee ad ogni istante temporale t. Il processo di computazione in tempo reale delle posizioni del veicolo (cioè, le osservazioni vengono prese al tempo t_k, e le posizioni risultanti vengono poi fornite allo stesso istante t_k) è detto *filtraggio*.

[15] Per una trattazione generale del filtro di Kalman si vada alle Appendici D ed E.

La computazione delle posizioni previste del veicolo agli istanti successivi t_k, basate sulle ultime misure a t_{k-1} è propriamente chiamata *previsione*, mentre la stima della posizione in cui il veicolo si trovava (ad esempio al tempo t_k), una volta che tutte le misure sono state post-processate al tempo t_{k+1}, è detta *arrotondamento*.

Anche se le tre procedure sono separate, e possono essere applicate indipendentemente, esse possono anche essere applicate sequenzialmente:

- Il **passo di previsione**: basato su *passate informazioni di posizionamento unitamente ad un modello cinematico*, viene computata la posizione prevista e la sua precisione alla successiva epoca di misura. Il modello cinematico è composto, come il modello di misura, di componenti funzionali (cioè, deterministiche) e stocastiche. Così questi *quattro* modelli devono essere considerati in base ad una particolare applicazione e ad un certo tipo di dati, il processo del filtro di progetto sta quindi nel selezionare i modelli appropriati.
- Il **passo di filtraggio (o di aggiustamento)**: a meno che non siano disponibili a priori buone stime dei parametri provenienti dal passo di previsione, filtrare i dati è un passaggio necessario. Fondamentalmente, i parametri stimati risultanti sono *combinazioni pesate delle quantità previste e dei dati di misura*. Il filtro di Kalman è una particolare forma generalizzata del filtro (metodo) dei minimi quadrati.
- Il **passo di arrotondamento**: in questo passo *tutte le misure vengono riprocessate*, a questo passo si arriva dopo che l'ultima misura è stata effettuata e che il passo di filtraggio è stato completato.

Come indicato sopra, l'implementazione del filtro richiede la specificazione dei modelli stocastici e matematici sia per il *sistema di misura* che per il *sistema dinamico*. Una volta che i modelli stocastici e matematici sono stati definiti, la loro implementazione nei i filtri di Kalman è, in linea di principio, relativamente semplice, sebbene vi siano varie implementazioni differenti che possono avere differenti vantaggi dal punto di vista computazionale, o per quanto riguarda la stabilità numerica o ancora per quanto concerne il controllo della qualità. Le tecniche basate sul filtro di Kalman sono particolarmente adatte per la navigazione GPS in quanto:

- La procedura standard dei minimi quadrati tratta ogni epoca di misura indipendentemente, e quindi le informazioni risultanti non possono essere impiegate nei sistemi dinamici, quale è ad esempio un veicolo in moto a cui un ricevitore GPS è attaccato.
- Permette la computazione rigorosa di misure precise ed affidabili esenti da errori di ellisse o da "errori marginalmente individuabili".
- Il filtro di Kalman è anche importante per i "controlli di qualità" o per procedure di "individuazione degli errori" che possono essere implementate in tempo reale per individuare "difetti" (dove dati di scarsa qualità vengono introdotti nel processo, o dove c'è un errore nella misura o nei modelli dinamici), bisogna allora *identificare* la fonte d'errore, ed *adattare* (o *aggiustare*) il sistema per assicurare che i risultati non siano errati a causa dei "difetti" di sistema.
- Stimare piccoli errori che affliggono i dati oltre varie epoche. Per esempio, molti errori di misura nelle moderne tecnologie di misura sembrano moti o spostamenti, che non sono apparenti al livello della singola epoca, perché essi appaiono come "rumore" nelle soluzioni standard epoca per epoca ai minimi quadrati.

- Tenendo conto delle informazioni del sistema dinamico, quali ad esempio il moto regolare del ricevitore GPS, è possibile ottenere le posizioni stimate persino se ci sono dati insufficienti, per esempio quando solo due satelliti sono visibili.
- E' un ottimo estimatore lineare in presenza di rumore bianco Gaussiano.
- Un filtro di Kalman può accettare dati come e quando essi sono stati misurati, i dati non devono essere "ridotti" a qualche epoca specifica.
- Il filtro di Kalman è adatto alla combinazione, o alla "fusione", di vari tipi di dati (inclusi quelli provenienti da sensori non GPS).

Alcuni ricevitori GPS incorporano filtri di Kalman come algoritmi di computazione navigazionale, ma il loro vero utilizzo è generalmente ovvio solo quando il sistema di posizionamento coinvolge vari sensori, ad esempio quando il GPS viene integrato con sensori "Dead Reckoning" (trad. posizionamento morto). Comunque, i filtri di Kalman non sono procedure "magiche" perché se i dati introdotti sono di qualità discutibile, o c'è un errore negli assunti che riguardano il modello del sistema dinamico, il posizionamento risulterà seriamente peggiorato.

4.9 Definizioni statistiche

Prima di parlare degli algoritmi OTF sperimentali ritengo sia corretto introdurre alcune definizioni statistiche:

a. **Variabile casuale (o stocastica)**: Mentre una variabile deterministica può assumere determinati valori con certezza, una variabile casuale assume determinati valori con una certa probabilità.
b. **Deviazione standard**: la d.s. è una misura della dispersione della distribuzione, ed ha le stesse unità di misura della variabile casuale analizzata:

$$\sqrt{Var(x)} = \sigma$$

noti la media e la deviazione standard della variabile si può definire il coefficiente di variazione: $C_V = \frac{\sigma}{\mu}$. Il coefficiente di variazione è un indicatore della dispersione della distribuzione.

c. **Varianza**: Si definisce v. (σ^2) della variabile casuale x il valore atteso della variabile $(x-\mu)^2$:

$$\sigma^2 = E[x-\mu]^2 \text{ dove:}$$

$$\mu_x = E(x) = \sum_{i=1}^{n} x_i P(x_i) \text{ se la variabile casuale è discreta}$$

$$\mu_x = E(x) = \int_{-\infty}^{\infty} x f(x) dx \text{ se la variabile casuale è continua}$$

la quantità μ_x o E(x) è detta media, o valore atteso, della variabile casuale x. La varianza è tanto più grande quanto maggiore è la dispersione dei valori di x attorno alla media μ. Risulta invece nulla se x assume un unico valore con probabilità 1.

d. **Covarianza**: La c. è la misura della varianza di una variabile rispetto ad un'altra ed è definita come:

$$Cov(x,y) = E[(x-\mu_x)(y-\mu_y)] =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f_{xy}(x,y)dxdy \quad \text{per variabili casuali discrete}$$

$$= \sum_i \sum_j (x-\mu_x)(y-\mu_y)P(x_i,y_j) \quad \text{per variabili casuali continue}$$

4.10 Algoritmi OTF sperimentali

Come già detto in merito all'OTF la ricerca è ancora ben lungi dal aver trovato un procedimento di (routine) applicabilità generale, a tale proposito in questo paragrafo verranno descritti alcuni algoritmi ancora in via di sperimentazione e verifica per la risoluzione dell'ambiguità di fase OTF. Il primo algoritmo è abbastanza classico, è infatti basato sul filtro di Kalman, il secondo si basa invece su un particolare modello ionosferico a due strati (il modello ionosferico che viene tipicamente utilizzato considera la ionosfera costituita da un unico strato bidimensionale, privo di spessore), il terzo ed ultimo algoritmo sceglie un unico set di interi sfruttando la cosiddetta *Fuzzy Logic* (ossia la Logica Sfumata) ed un procedimento decisionale MADM.

4.10.1 Algoritmo OTF basato sul filtro di Kalman per applicazioni cinematiche con lunghi baseline

È sempre stata una grande sfida determinare e fissare le ambiguità di fase GPS quando devono considerarsi lunghi *baseline*. La sfida è ancora più grande per applicazioni GPS cinematiche, generalmente la difficoltà nel risolvere le ambiguità è dovuta alla decorrelazione dei *bias* nelle osservazioni GPS. Come è ben noto, le osservazioni GPS alle stazioni base e *rover* sono influenzate da differenti effetti atmosferici ed errori d'orbita dei satelliti man mano che la distanza tra le stazioni aumenta. Quando le misure di codice sono incorporate alle misure di fase, il *multipath* può essere la fonte d'errore dominante che rende difficile la risoluzione delle ambiguità a causa del suo comportamento quasi-casuale oltre un tempo misurato relativamente breve. Nelle situazioni cinematiche, non è facile modellare il rumore di osservazione in quanto il moto delle piattaforme può mascherare alcuni aspetti del rumore di osservazione. Usualmente il rumore può essere ben modellato statisticamente tramite una funzione la cui variabile indipendente è l'angolo di elevazione.
Per ottenere soluzioni ottimali con il metodo dei minimi quadrati, un modello funzionale (o deterministico) ed un modello stocastico devono essere specificati correttamente. Il modello funzionale descrive la relazione tra le osservazioni ed i parametri incogniti mentre il modello stocastico rappresenta le caratteristiche del rumore delle osservazioni. Attualmente, la sfida che ci troviamo di fronte nelle applicazioni cinematiche con lunghi *baseline* consiste nella corretta specificazione dei modelli senza ignorare i problemi sopra menzionati.
In questo caso, il problema legato al modello funzionale è che il numero di parametri incogniti è più grande del numero di osservazioni, anche se cerchiamo di ottenere osservazioni esterne, quali ad esempio quelle ricavabili da un orologio esterno. Oltretutto,

risulta essere molto complicato specificare un modello stocastico corretto se noi optiamo per un modello funzionale semplificato che ignora certi parametri, a causa degli effetti residui di tali parametri e del moto delle piattaforme. Come programma fondamentale nel trattamento delle osservazioni GPS, noi dobbiamo anche considerare un controllo della qualità degli risultati; ad esempio, come noi possiamo implementare una routine robusta di trattamento dei *cycle slip*? Specialmente per applicazioni cinematiche con lunghi *baseline*.
Fondamentalmente, l'approccio che stiamo per descrivere consiste nel tentativo di superare queste sfide seguendo una procedura generalizzata che consiste nel tenere conto dei problemi citati. Il prototipo dell'approccio è basato nel caso in cui osservazioni GPS a doppia frequenza sono disponibili. Nelle situazioni in cui osservazioni esterne sono altresì disponibili, l'approccio può integrare le informazioni aggiuntive senza complessità immotivate.

Considerazioni per un approccio affidabile

Come risulta dall'esperienza, il modello stocastico è tipicamente più difficile da trattare del modello funzionale se consideriamo un approccio affidabile per applicazioni cinematiche con lunghi *baseline*. Assumendo che il modello funzionale includa tutti i parametri incogniti significativi (ad esempio, quelli associati con gli effetti atmosferici, gli errori d'orbita, il *multipath*, e via dicendo) eccetto il rumore di sistema del ricevitore (cioè: il rumore dell'antenna e le perdite dei cavi), noi possiamo considerare più facilmente il modello stocastico. In questo caso, i problemi associati al modello stocastico sono:

- correlazione incrociata (tra differenti tipi di osservazione);
- correlazione temporale (tra diverse epoche);
- correlazione spaziale (tra diversi canali);
- dipendenza dall'angolo di elevazione;
- distribuzione probabilistica dell'errore.

In pratica noi consideriamo che tutti i parametri descriventi i problemi di modellazione stocastica possono essere calibrati in laboratorio.
Se il modello funzionale è corretto, tali parametri possono essere usati sul sito remoto senza calibrazione. Comunque, si nota che la dipendenza dall'angolo di elevazione del rumore di sistema spesso varia con la particolare situazione cinematica. La dipendenza dall'angolo di elevazione del sistema è indotta principalmente dal modello di acquisizione del dato dell'antenna del ricevitore, insieme ad altri fattori quali l'attenuazione atmosferica del segnale. L'angolo di elevazione è normalmente computato rispetto all'orizzonte del piano geodetico locale ed al centro di fase dell'antenna, senza considerare però l'orientamento della stessa.
Conseguentemente, la relazione tra l'acquisizione dell'antenna e l'angolo di elevazione del segnale può con difficoltà considerare il cambio di orientamento dell'antenna, come spesso capita nelle situazioni cinematiche. Se noi usiamo un modello funzionale che include tutti i parametri incogniti significativi, ci troviamo di fronte un problema risolubile con il metodo dei minimi quadrati o con il filtro di Kalman.
Sebbene questo problema possa essere trattato in seconda istanza con un metodo di trasformazione dei parametri per ridurne il loro numero, una difficoltà inerente permane ancora: i parametri stimati possono essere affetti da errore (scintillazione ionosferica, scosse, etc.) perché le componenti ad altissima frequenza dei parametri incogniti non possono essere stimate in condizioni di dati insufficienti.

Utilizzando un filtro di Kalman, noi possiamo aumentare il più alto ordine temporale che deriva dal vettore di stato, dato dai parametri incogniti in modo da tener conto delle componenti ad altissima frequenza. Comunque, gli stati stimati possono ancora essere affetti da errori.
Nel contesto del trattamento del segnale, questo problema è legato al tasso di campionamento (ad esempio, un tasso di campionamento di 1 Hz), le serie temporali possono contenere informazioni ad una larghezza di banda di 0,5 Hz. Per ottenere una larghezza di banda più ampia che includa componenti ad altissima frequenza, è necessario aumentare il tasso di campionamento abbastanza da rimuovere l'effetto dei falsi dati, in modo da non influenzare i parametri stimati.

Risoluzione dell'ambiguità di fase ad interi tramite un processo di ricerca basato sul filtro di Kalman

Vari possono essere gli approcci al problema: "quale strategia dovrebbe essere preferibile per trattare le osservazioni GPS?" In termini di implementazione, la nostra risposta è: "un approccio" basato sul filtro di Kalman combinato con un metodo di ricerca dell'ambiguità che può trattare con i modelli funzionale e stocastico in modo ottimale.

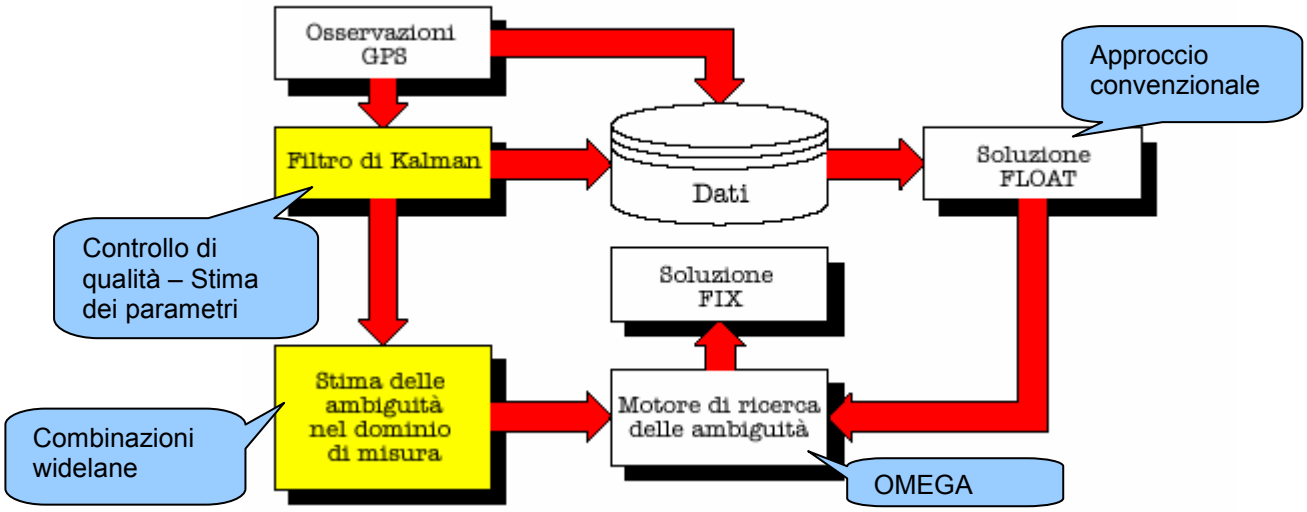


Diagramma funzionale a blocchi per il fissaggio ad interi dell'ambiguità di fase nelle applicazioni GPS cinematiche con lunghi baseline

Approccio con il filtro di Kalman

Si è visto che un approccio basato sul filtro di Kalman può efficientemente implementare schemi di controllo della qualità quali trattamento dei *cycle slip* (cioè individuazione, identificazione ed adattamento), i parametri stimati del filtro possono essere usati in seconda istanza nel processo di ricerca dell'ambiguità cosicché i parametri stimati non sono affetti da errore. Comunque, i concetti fondamentali legati a questa implementazione sono:

1. Come possiamo ridurre il numero di parametri incogniti nel vettore di stato del filtro?
2. Come possiamo assicurare l'osservabilità del modello di sistema in condizioni di dati insufficienti?
3. Quale metodo di implementazione è più efficiente?

Sostanzialmente, il problema è che il numero dei parametri incogniti è molto maggiore di quello delle osservazioni. Questo è un problema inerente alle applicazioni che sfruttano le misure di fase e diventa un problema sostanziale in approccio come il nostro che prova a stimare tutti i parametri di errore (tutti eccetto il *multipath*). Per ridurre il numero di parametri incogniti, lo schema delle doppie differenze viene utilizzato. In aggiunta, misure di fase a doppia frequenza (L1 ed L2) e misure di codice (P1 e P2, oppure C/A e P2) sono utilizzate per aumentare la ridondanza delle osservazioni, ed i parametri incogniti sono trasformati per assicurare l'osservabilità del dato modello di sistema. Un filtro di Kalman separato è stato implementato per ogni serie temporale di doppie differenze perché la sua programmazione e modellazione stocastica è semplice. Come risultato otteniamo il seguente vettore di stato:

$$\hat{x} = \begin{bmatrix} \hat{L} \ldots \hat{I} \ldots \hat{B}_1 \ldots \hat{B}_2 \ldots \hat{n}_1 \, \hat{n}_2 \end{bmatrix}^T$$

Dove:

$$\begin{aligned}
\hat{L} &= L + \frac{1}{\gamma-1}\left(\gamma b_1 - b_2\right) + \frac{1}{\gamma-1}\left(\gamma B_1^{'0} - B_2^{'0}\right) \\
\hat{I} &= I + \frac{1}{\gamma-1}\left(b_1 - b_2\right) - \frac{1}{\gamma-1}\left(B_1^{'0} - B_2^{'0}\right) \\
\hat{B}_1 &= B_1^{'} - B_1^{'0} \\
\hat{B}_2 &= B_2^{'} - B_2^{'0} \\
\hat{n}_1 &= n_1 + \frac{1}{\gamma-1}\left[-\left(\gamma+1\right)B_1^{'0} + 2B_2^{'0}\right] \\
\hat{n}_2 &= n_2 + \frac{1}{\gamma-1}\left[-2\gamma B_1^{'0} + \left(\gamma+1\right)B_2^{'0}\right]
\end{aligned}$$

ed anche:

$$\begin{aligned}
B_1^{'} &= B_1 + \frac{1}{\gamma-1}\left[-\left(\gamma+1\right)b_1 + 2b_2\right] \\
B_2^{'} &= B_2 + \frac{1}{\gamma-1}\left[-2\gamma b_1 + \left(\gamma+1\right)b_2\right] \\
B_1^{'0} &= B_1^0 + \frac{1}{\gamma-1}\left[-\left(\gamma+1\right)b_1^0 + 2b_2^0\right] \\
B_2^{'0} &= B_2^0 + \frac{1}{\gamma-1}\left[-2\gamma b_1^0 + \left(\gamma+1\right)b_2^0\right]
\end{aligned} \qquad (0)$$

Dove L sta per effetto combinato di un *range* geometrico assunto a priori, gli errori d'orbita dei satelliti e il ritardo troposferico ($L=\rho+s+\tau$); I è il ritardo ionosferico; b è il *multipath* nelle misure di fase; B il *multipath* nelle misure di codice; n per le ambiguità (espresse in metri);

"…" sono gli ordini temporali superiori derivanti dai parametri; $\gamma=(\lambda_2/\lambda_1)^2$ è una costante; lo zero all'apice sta per il valore iniziale dell'errore (all'inizio delle osservazioni); gli "1" e "2" al pedice corrispondono rispettivamente ad L1 ed L2, P1 e P2 (o C/A e P2). Il simbolo "^" indica un parametro trasformato.
Si può notare che ogni parametro trasformato del precedente gruppo di equazioni include un valore di parametro vero, il *multipath* sulle misure di fase di L1 ed L2, il *multipath* iniziale costante ($B_1^{'0}$ e $B_2^{'0}$), e il rumore di sistema del ricevitore.
Il *multipath* iniziale costante può essere separato dai parametri stimati durante il processo di ricerca dell'ambiguità, ma il *multipath* nelle misure di fase è molto più difficile da stimare in questo approccio cosicchè non possiamo utilizzare osservazioni addizionali quali il tasso di rumore del segnale (SNR *Signal to Noise Ratio*) per l'osservazioni basate su misure di fase. Si può notare che in qualche caso i parametri stimati (determinati in questo modo) possono essere affetti da errori. Noi non possiamo avere parametri stimati senza errori se i *cycle slip* non sono stati trattati perfettamente.

Controllo di qualità

Finché non consideriamo un *cycle slip* come parametro incognito nel modello funzionale, noi dobbiamo individuarlo e rimuoverlo dalle osservazioni. Se non riusciamo a farlo, i parametri stimati con il filtro di Kalman saranno sempre affetti da errori. Fondamentalmente, noi abbiamo usato una procedura per il trattamento dei *cycle slip* basata sulla previsione dei residui con il filtro di Kalman, ed abbiamo riscontrato che non funziona come ci si aspetta nelle situazioni cinematiche. Questo problema è dovuto al moto delle piattaforme ed eventualmente al tasso di campionamento.
Per rafforzare la procedura su questi fronti, noi utilizziamo una tecnica di mascheramento basata sull'intersezione logica di condizioni necessarie e sufficienti per l'individuazione e identificazione dei *cycle slip*. Quando si verifica un *cycle slip*, noi possiamo notare un impulso nel grafico delle serie temporali delle quadruple differenze (ottenute differenziando osservazioni consecutive alle triple differenze). Questo consente di ottenere una condizione necessaria per l'identificazione dei *cycle slip*. Come approccio convenzionale incorporato nel filtro di Kalman, noi possiamo utilizzare la previsione dei residui per individuare i *cycle slip*.
Comunque, bisognerebbe applicare con molta cautela questo approccio perché la previsione dei residui è molto sensibile al moto delle piattaforme ed al tasso di campionamento delle osservazioni. Un altro approccio consiste nel usare il ritardo ionosferico stimato, ma anche stavolta bisogna muoversi con cautela perché ci sono dei casi in cui i *cycle slip* non possono essere individuati, ad esempio quando abbiamo simultaneamente *cycle slip* della stessa estensione (espressa in metri), e quando poi si verifica il caso in cui ci sono *cycle slip* su entrambe le portanti che cancellano ogni altra interruzione nelle combinazioni IONO-FREE (ad esempio $\frac{1}{77}c_1 - \frac{1}{60}c_2 = 0,$ dove c_1 e c_2 rappresentano i *cycle slip* su L1 ed L2 espressi in numero di cicli). Ciononostante, in un senso molto ampio, questi due approcci – previsione dei residui e ritardi ionosferici stimati - consentono di ottenere condizioni sufficienti per individuale i *cycle slip*.
Così, noi abbiamo trovato che questa procedura è quasi perfetta per quanto concerne l'individuazione e l'identificazione dei *cycle slip*. Mentre per il loro adattamento (cioè per la rimozione dalle osservazioni) dovrebbe essere eseguita con molta cautela, perché prima della rimozione l'estensione dei *cycle slip* deve essere determinata correttamente. Se noi proviamo a determinare l'estensione di un *cycle slip* utilizzando la previsione dei residui

del filtro di Kalman, noi possiamo introdurre un nuovo errore nelle osservazioni. Una semplice strategia per evitare questo problema consiste nel resettare il vettore di stato del filtro di Kalman ogni volta che un *cycle slip* viene individuato ed identificato.

Stima del rumore di sistema del ricevitore

Se il modello stocastico non è corretto, questo influenzerà la computazione dei parametri stimati di qualche grado. Come già detto, è meglio non utilizzare una funzione la cui variabile indipendente sia l'angolo di elevazione per il modello stocastico, nella raccolta dei dati provenienti dalle piattaforme in moto. Un approccio alternativo e potenzialmente migliore, può essere derivato direttamente da misurazioni della qualità di ogni osservazione basata su misure di fase. Questa informazione è contenuta nel SNR (o alternativamente nel tasso di densità e potenza del rumore della portante, C/N_0). Sebbene i meriti potenziali nel uso delle informazioni SNR, come schema di modellazione stocastica sia stato già discusso, un esame comprensivo della tecnica è stato effettuato solo recentemente. Anche se alcuni produttori di ricevitori GPS hanno incluso i valori SNR nei i loro flussi di dati, valori di SNR significativi non sono facili da acquisire.
Il seguente concetto rappresenta un altro approccio alternativo che può essere derivato direttamente dalle serie temporali delle osservazioni. Noi possiamo usare serie temporali di quintuple differenze (ottenute differenziando osservazioni consecutive di quadruple differenze dopo aver cancellato gli impulsi dei *cycle slip*) per stimare il rumore di sistema del ricevitore per osservazioni condotte in modalità cinematica. Noi abbiamo scelto questo approccio perché le serie temporali delle quadruple differenze sono già state ottenute per il controllo di qualità come descritto precedentemente. Ciononostante, questo approccio può essere implementato senza complessità immotivata, noi assumiamo che gli effetti dei parametri incogniti (eccetto il rumore di sistema del ricevitore) siano stati rimossi nelle serie temporali delle quintuple differenze. Ad esempio, consideriamo le serie temporali delle quintuple differenze di fase:

$$\dddot{\Phi}_1 = \dddot{\rho} + \dddot{\tau} + \dddot{s} - \dddot{I} + \dddot{b}_1 + \dddot{n}_1 + \dddot{\varepsilon}_1 \tag{1}$$

dove Φ_1 è l'osservabile L1 delle doppie differenze. Usando una serie di Taylor mono dimensionale includendo infinitesimi temporali di ordine superiore per ogni parametro incognito, si ha:

$$S(t) = S(t_0) + S'(t_0)(t - t_0) + \frac{1}{2}S''(t_0)(t - t_0)^2 + \frac{1}{6}S'''(t_0)(t - t_0)^3 + R(t) \tag{2}$$

dove S rappresenta ogni parametro incognito ed R è il resto di Lagrange. Assumendo che l'intervallo temporale di osservazione ($t-t_0$) è lo stesso di $\ddot{A}$ per le serie temporali, otterremo la seguente quintupla differenza:

$$\dddot{S}(t_3) = S(t_3) - 3S(t_2) + 3S(t_1) - S(t_0) = S'''(t_0)\Delta^3 + \Sigma_R(t_3), \tag{3}$$

dove $\Sigma_R(t_3)$ è la quintupla differenza per il resto R. Sostituendo l'equazione (3) si avrà:

$$\dddot{\Phi}_1(t_3) = \left[\sum_{\forall S} S'''(t_0)\right]\Delta^3 + \sum_{\forall S}\left[\Sigma_R(t_3)\right] + \dddot{\varepsilon}_1(t_3) \tag{4}$$

dove:

$$\sum_{\forall S} S'''(t_0) = \lfloor \rho''' + \tau''' + s''' - I''' + b_1''' + n_1''' \rfloor (t_0). \qquad (5)$$

se l'effetto dei termini a destra dell'equazione (5) è piccolo abbastanza da essere trascurabile e/o il tasso di campionamento $\left(1/\ddot{A}\right)$ è alto nell'equazione (4), e l'effetto del secondo termine a destra dell'equazione (4) (ad esempio, l'effetto della quintupla differenza per il resto R) è anche piccolo abbastanza da essere trascurabile, noi possiamo considerare un'inferenza accettabile:

$$\dddot{\Phi}_1 \approx \dddot{\varepsilon}_1 \leftarrow \text{inferenza}$$

Processo di ricerca dell'ambiguità

Usando le stime del vettore di stato, noi possiamo trasformare le osservazioni originali delle doppie differenze di fase in quelle da utilizzare per il processo di ricerca dell'ambiguità. Lo scopo di questa trasformazione è di ridurre il numero di parametri incogniti nel passo di ricerca dell'ambiguità. Comunque, ci possono essere alcuni svantaggi nell'uso di questa trasformazione, ad esempio il rumore di sistema del ricevitore è aumentato e risulta essere correlato al tempo, noi usiamo una trasformazione *iono-free* per ridurre questi svantaggi:

$$\Phi_1 + \hat{I} = L + \frac{1}{\gamma - 1}\left(\gamma b_1 - b_2\right) + n_1 - \frac{1}{\gamma - 1}\left(\gamma B_1^{'0} - B_2^{'0}\right) + \varepsilon_1^{'}$$

osservabile alle doppie differenze con combinazione *iono-free*

$$\Phi_2 + \gamma \hat{I} = L + \frac{1}{\gamma - 1}\left(\gamma b_1 - b_2\right) + n_2 - \frac{1}{\gamma - 1}\left(\gamma B_1^{'0} - B_2^{'0}\right) + \varepsilon_2^{'}$$

E' vero che le osservazioni trasformate sono simili alla combinazione lineare *iono-free* ma hanno rumore d'osservazione meno intenso. Il rumore d'osservazione correlato al tempo può essere stimato utilizzando la matrice di varianza covarianza che è ottenuta adattivamente dal filtro di Kalman. Utilizzando l'osservazione trasformata per ogni osservazione alle doppie differenze noi abbiamo aumentato il numero di equazioni d'osservazione:

$$\ddot{O}_1 + \hat{I} - \left(\tilde{n}_0 + \hat{o}_0\right) = Ax + \ddot{E}_1 N_1 + s + \frac{1}{\gamma - 1}\left(\gamma b_1 - b_2\right) - \frac{1}{\gamma - 1} B_{12}^{'0} + \overset{\circ}{a}_1^{'}$$

$$\ddot{O}_2 + \hat{\gamma I} - \left(\tilde{n}_0 + \hat{o}_0\right) = Ax + \ddot{E}_2 N_2 + s + \frac{1}{\gamma - 1}\left(\gamma b_1 - b_2\right) - \frac{1}{\gamma - 1} B_{12}^{'0} + \overset{\circ}{a}_2^{'}$$

equazioni alle doppie differenze trasformate

dove:

- x include le componenti incognite del *baseline* e il ritardo *troposferico* residuo;
- A è la corrispondente matrice disegno (contiene solo coeff. 0, 1, -1) per x;
- N è il vettore ambiguità (espresso in unità di ciclo);
- $\ddot{E}$ è la corrispondente matrice disegno per N;
- s è il vettore degli errori d'orbita (trascurabile);
- b è il vettore del *multipath* nelle misure di fase (trascurabile);
- $B_{12}^{'0} = B_1^{'0} - B_2^{'0}$ è un vettore costante;
- $\tilde{n}_0$ e $\hat{o}_0$ sono, rispettivamente, le stime iniziali del *range* geometrico e del ritardo troposferico.

Nell'implementazione pratica delle precedenti equazioni, noi assumiamo che il *multipath* nelle misure di fase sia trascurabile e che dati di orbita precisi siano disponibili. Come abbiamo capito osservando il precedente gruppo di equazioni le ambiguità non possono essere separati dai parametri $B_{12}^{'0}$ perché essi sono costanti.
Questo problema può essere risolto utilizzando una combinazione *widelane* delle stime $\hat{n}_1$ ed $\hat{n}_2$ espresse in metri. Dal gruppo di equazioni (0) noi possiamo formulare una combinazione *widelane* come:

$$\hat{n}_1 - \hat{n}_2 = \left(\Lambda_1 N_1 - \Lambda_2 N_2\right) - B_{12}^{'0}$$

Quando il filtro di Kalman converge, la combinazione *widelane* tende velocemente ad una costante. Questo significa che i parametri $B_{12}^{'0}$ possono essere determinati nel processo di ricerca dell'ambiguità dove N_1 ed N_2 sono dati come valori noti, e la combinazione *widelane* $\hat{n}_1 - \hat{n}_2$, converge.
Per il processo di ricerca, noi usiamo un approccio indipendente di ricerca dell'ambiguità, finché restano quattro parametri incogniti nell'equazione d'osservazione dopo che le osservazioni sono state trasformate, rimangono sempre otto livelli di ricerca (rispettivamente quattro livelli di ricerca per N_1 ed N_2, che sono le ambiguità di L1 ed L2) senza considerare il numero d'osservazioni alle doppie differenze. In questo caso, lo spazio di ricerca potrebbe essere enorme persino se un piccolo spazio di ricerca viene considerato.
Questo significa che il processo di ricerca dell'ambiguità potrebbe essere esoso in termini di tempo il che non è appropriato per un sistema in tempo reale. Allo scopo di superare questo problema, viene utilizzato un motore di ricerca efficiente dell'ambiguità, noto con il nome di OMEGA (*Optimal Method for Estimating GPS Ambiguities*, trad. Metodo Ottimale per la Stima delle Ambiguità GPS).

4.10.2 Risoluzione dell'ambiguità di fase On the Fly a piu' di 100 km dalla piu' vicina stazione di riferimento, con l'aiuto della tomografia ionosferica

Consideriamo adesso il problema di come risolvere l'ambiguità di fase delle portanti GPS in modo preciso e veloce, quando il *rover* si trova a più di 100 km dalla più vicina stazione di riferimento, in modo da ottenere posizioni in tempo reale (r.m.s.) e con precisione sub-decimetrica. Per far questo in modo affidabile, la rifrazione ionosferica deve risolta molto accuratamente. Per distanze superiori a poche centinaia di chilometri, i dati GPS in doppia frequenza provenienti dalle stazioni di controllo possono essere usati per ottenere le necessarie informazioni sulla rifrazione ionosferica. La ionosfera sopra la zona servita dalle stazioni deve essere mappata utilizzando i dati delle misure di fase, anche in tempo reale, con l'aiuto della tomografia ionosferica. La mappa risultante è utilizzata per produrre informazioni che vengono trasmesse all'utente, lungo le correzioni di *range* e di tempo. L'utente può allora calcolare correzioni molto precise per la rifrazione ionosferica nella posizione del ricevitore *rover*.

Introduzione

Utilizzando quanto è noto dal WAAS (*Wide Area Augmentation System* trad. Sistema di Incremento per zone estese), i dati *pseudorange* del GPS vengono utilizzati per posizionare in tempo reale gli aerei, con precisione metrica, relativamente a stazioni di riferimento distanti centinaia di chilometri. Se le misure di fase sono utilizzate al posto di quelle di codice si può raggiungere un livello di precisione sub-decimetrico. Comunque, le misure di fase sono ambigue, e la risoluzione esatta delle loro ambiguità è ostacolata dalla rifrazione ionosferica che si verifica già lungo poche decine di chilometri. Questa distanza diminuirà quando l'attività ionosferica aumenterà in seguito al raggiungimento del Massimo Solare che probabilmente si verificherà nei primi anni del 2000. Come si può vedere dalla figura 1 in basso, successivamente ci si aspetta che l'attività solare rimanga alta per parecchi anni.

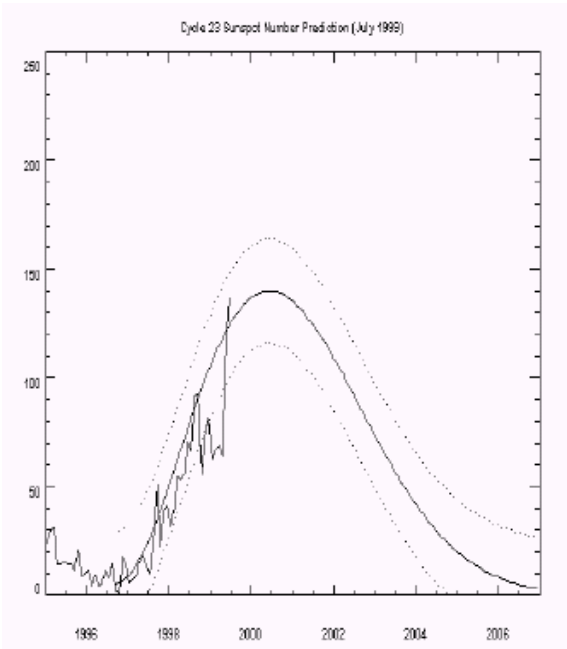


Figura 1. Numero di macchie solari in rapporto al ciclo solare attuale previsto e misurato

Un modo per risolvere l'ambiguità di fase On the Fly (OTF) oltre distanze di centinaia di chilometri tra le stazioni *rover* e fisse, è di fare correzioni ionosferiche alle doppie differenze sui dati del *rover*. Correzioni molto precise possono essere effettuate sia in tempo reale che in modalità post, utilizzando informazioni ionosferiche fornite da una rete di stazioni di controllo permanenti con ricevitori a doppia frequenza. Le correzioni possono essere basate su diversi tipi di modelli ionosferici, noi abbiamo utilizzato, in modalità tempo reale, un modello tomografico a due strati. Questo modello ha dimostrato di fornire stime accurate e veloci del Contenuto Elettronico Totale (TEC, Total Electron Content) in particolare nelle regioni vicine all'equatore geomagnetico o in periodi, come il Massimo Solare, in cui vi è una elevata variabilità della densità elettronica.

Effettuazione delle correzioni ionosferiche

La figura 2 in basso mostra l'ordine dei passi eseguiti per ottenere correzioni ionosferiche alle doppie differenze per il ricevitore *rover*.

Tre passi principali possono essere enfatizzati.

(a) Il modello ionosferico è stato creato utilizzando misure di fase provenienti dalle stazioni di riferimento. La ionosfera è stata poi suddivisa in celle tridimensionali in un sistema di riferimento fisso chiamato: "latitudine tempo/geodetica" (celle di superficie 5x2 gradi ed altezza di 60-740-1240 km) in tali celle si è assunto che la densità elettronica durante il passaggio dei segnali GPS sia costante.
Dove $L1=\lambda_1\ \varphi_1$ (λ_1 è la lunghezza d'onda e φ_1 è la fase corrispondente alla frequenza L1), ed $L2=\lambda_2\ \varphi_2$ (per la frequenza L2).

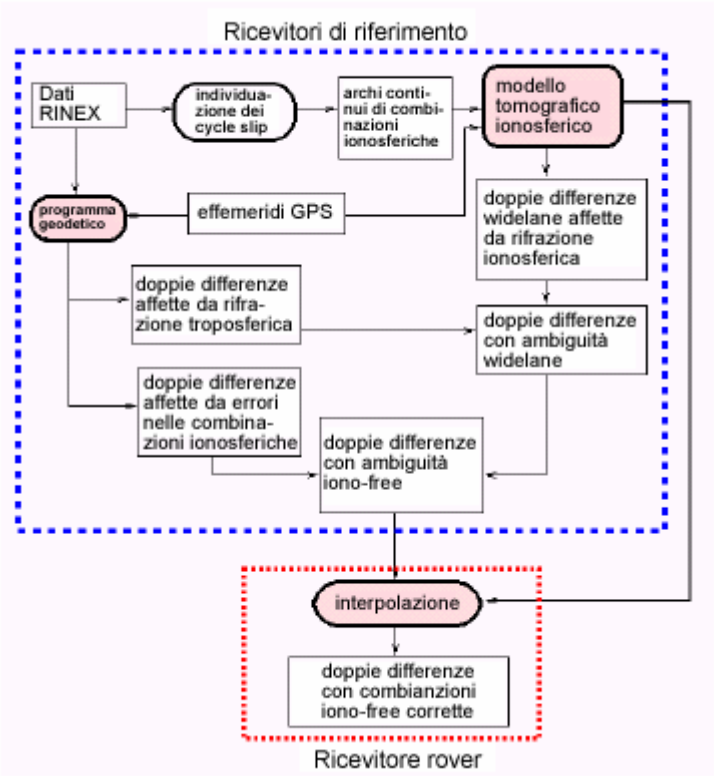


Figura 2. Schema della procedura di correzione ionosferica

Allora, per ogni coppia satellite-ricevitore, per la dipendenza della frequenza dalla propagazione dei ritardi delle onde radio, dovuti alla ionosfera, valgono le seguenti equazioni per le combinazioni ionosferiche di fase LI=L1-L2 (proporzionale allo *Slant TEC*, o STEC trad. TEC deviato):

$$LI \approx \sum_i \sum_j \sum_k (N_e)_{i,j.k} \Delta s_{i,j.k} + \lambda_1 b_1 - \lambda_2 b_2$$

dove i,j,k sono gli indici corrispondenti al tempo locale (ad ogni istante temporale la densità elettronica cambia), alla latitudine geodetica e all'altezza, $(N_e)_{i,j,k}$ è la corrispondente densità elettronica libera, $\Delta s_{i,j.k}$ è la lunghezza della traiettoria dei raggi che incrociano la cella i,j,k ($\Delta s_{i,j.k} = 0$ per le celle "oscure" che schermano completamente i segnali GPS, ne impediscono cioè il passaggio causando *cycle slip*). Ed infine, b_1 e b_2 sono i termini di ambiguità, associati alle lunghezze d'onda λ_1 e λ_2, includenti anche i ritardi strumentali.
Equazioni simili possono essere scritte per le osservazioni ionosferiche PI=P2-P1, basate sulle misure di codice P1 e P2. Si verifica però che i ritardi strumentali dei codici del trasmettitore e del ricevitore D^T e D_R appaiono al posto dei termini di ambiguità.
Stimare la densità elettronica $(N_e)_{i,j,k}$ dalle misurazioni di doppia frequenza consiste nel risolvere il problema opposto, i parametri del modello a due strati vengono ottenuti utilizzando un filtro di Kalman con intervalli di acquisizione di 10 minuti. Si verifica cioè che nel corso della giornata, ad intervalli di 10 minuti, venga stimato il TEC di ogni cella istante per istante. Si assumendo il seguente comportamento stocastico per il trattamento del rumore:

- *Densità elettronica N_e*: percorso casuale con densità spettrale pari alla derivata temporale della portata elettronica: $\frac{dQ}{dt} = 10^8$ elettroni/m^3/ora (dimensionalmente si ha: $\left[\frac{n° elettr. \times tempo}{volume}\right]$ che è la densità elettronica di volume per un tempo), nel suddetto sistema di riferimento solare fisso.
- *Ritardi strumentali D^T e D_R*: constanti (in realtà tendono a peggiorare con l'uso).
- *Ambiguità $\lambda_1 b_1$-$\lambda_2 b_2$*: constanti lungo ogni arco di fase continua, ed assenti durante i *cycle slip*.

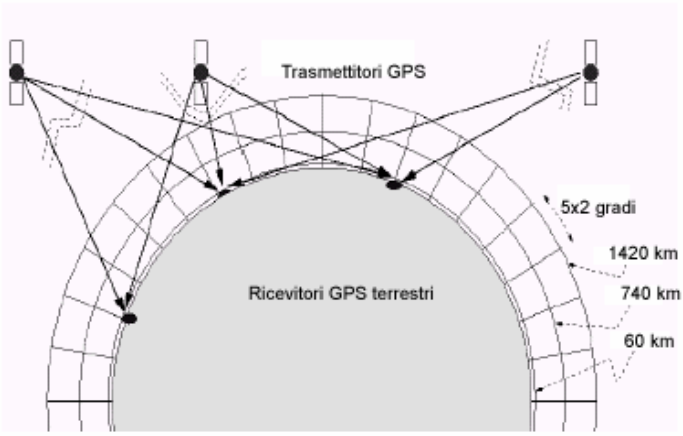


Figura 3. Schema del modello tomografico adottato per stimare il contenuto elettronico dalle stazioni di riferimento a terra

La densità elettronica nel modello può essere stimata solo con le misure di fase, ma l'introduzione delle misure di codice (con un peso appropriato $\sigma_{P1} = 100\sigma_{L1}$) fornisce più forza nella stima delle ambiguità di fase, e consente la stima dei ritardi strumentali.
Dopo l'inizializzazione del filtro, la soluzione è principalmente guidata dalla misure di fase, e quindi praticamente immune all'*Anti-Spoofing* e al *multipath* di codice. Questo è un miglioramento ulteriore su metodi che usano fasi pre-allineate con il codice, o codici arrotondanti.
(b) Trovare i tratti ionosferici non ambigui nei siti di riferimento, con l'aiuto di un programma geodetico. Le coordinate delle stazioni di controllo permanenti sono già note al livello centimetrico, in un sistema di riferimento ben definito si può usare un programma geodetico di analisi dei dati GPS (in questo caso il GIPSY) per stimare, in tempo reale e con precisione centimetrica, gli errori della combinazione *iono-free* (L_C) e la rifrazione troposferica residua rimasta dopo la correzione con un modello atmosferico standard. E' ora possibile computare *range* geometrici accurati tra stazioni e satelliti corretti per la troposfera e la ionosfera (utilizzando il modello ionosferico). Per distanze di poche centinaia di chilometri, gli errori nelle effemeridi trasmesse possono essere tranquillamente ignorati.
Da quando le stazioni di controllo operano continuamente, è anche possibile stimare continuamente il modello ionosferico, l'errore b_{Lc} di L_C e la rifrazione troposferica. Quando un ricevitore *rover* inizia ad operare, molte di queste quantità date dalle stazioni di controllo dovrebbero avere stime sequenziali ben convertite pronte per l'uso immediato.

Le ambiguità *widelane* sono state trovate arrotondando le differenze tra le *widelane* e i *range* geometrici corretti dalla rifrazione. E' molto importante che l'errore combinato del *range* e il rumore della *widelane* sia meno di 43 cm, o metà della lunghezza d'onda *widelane*. Un importante fattore che limita l'accuratezza delle correzioni ionosferiche è la bassa elevazione dei satelliti, alcuni esperimenti hanno dimostrato che tali correzioni sono tranquillamente utilizzabili da 20° di elevazione in su.
Dagli errori stimati al livello centimetrico dalle combinazioni *iono-free* b_{Lc} e dalle ambiguità *widelane* risolte N_w (alle doppie differenze) e possibile ottenere le ambiguità N1 ed N2, di L1 ed L2.
Le equazioni rilevanti sono:

$$b_{Lc} = \frac{1}{2}\left[\lambda_\omega N_w + \lambda_\upsilon (N1 + N2)\right], \text{ e così:}$$

$$N1 + N2 = \text{Intero più vicino}\left(\frac{2b_{Lc} - \lambda_\omega N_w}{\lambda_\upsilon}\right)$$

$$N1 = 0.5\left[N_w + (N1 + N2)\right]$$

$$N2 = N1 - N_w,$$

dove $\lambda_\omega \approx 86$ cm, $\lambda_v \approx 11$ cm, sono rispettivamente le lunghezze d'onda delle combinazioni *widelane* e *narrowlane*.
Finalmente, abbiamo trovato il contenuto elettronico totale deviato non ambiguo alle doppie differenze STEC: $LI=L1-L2-(\lambda_1 N1- \lambda_2 N2)$ per le stazioni di controllo.

(c) Interpolazione[16]: allo scopo di trovare le correzioni ionosferiche per il *rover*, due schemi sono stati considerati per interpolare il contenuto LI non ambiguo alle doppie differenze con i valori attuali delle stazioni:

- Interpolazione lineare delle correzioni ionosferiche non ambigue.
- Se invece le correzioni sono ambigue si procede con l'interpolazione guidata dal modello ionosferico stesso: la deviazione della previsione del modello dalla rifrazione ionosferica, non ambigua, in tutte le stazioni di riferimento è interpolata linearmente.

Risoluzione delle ambiguità On the Fly

Due passi principali sono richiesti nell'algoritmo utilizzato in questo studio.

(a) Risoluzione widelane. La combinazione *widelane* alle doppie differenze viene prima corretta per la ionosfera con il modello-interpolato di rifrazione, in tal modo l'errore della combinazione *iono-free* L_C viene sottratto. Finalmente, una stima dell'errore in L_C viene sottratta, tutto ciò si ripete per ogni doppia differenza in cui c'è una correzione ionosferica affidabile (satelliti sopra i 20° di elevazione).
Il risultato di questa operazione è l'ambiguità *widelane* (espressa in metri), più il rumore delle misure di fase, meno l'errore nella correzione ionosferica[17] e nel L_C stimato:

$$\lambda_w N_w + \text{rumore(widelane)} - \text{errore(ionosferico} + \overbrace{\text{errore } L_C}^{b_{Lc}}) =$$
$$= \text{widelane} - L_C - \text{corr.ion.} - \text{err. } L_C \text{ stimato}$$

L'ambiguità *widelane*, N_w, può essere trovata arrotondando questo risultato al più vicino numero intero di lunghezze d'onda *widelane*. Gli errori nella rifrazione troposferica computata, nelle coordinate della stazione di riferimento, e nelle effemeridi dei satelliti vengono cancellati sottraendo L_C dalle *widelane*. Assumendo che la correzione ionosferica sia sufficientemente accurata da risolvere le ambiguità L1 ed L2 (l'errore è inferiore ad 2.7 cm in LI), l'incertezza principale è nel errore L_C. Assumendo inoltre che le principali fonti d'incertezza siano distribuite normalmente (così dicendo ci riferiamo alla distribuzione di probabilità normale, cioè alla campana di Gauss) l'errore nel L_C dovrebbe essere inferiore ad ¼ di un *widelane* (< 21.5 cm), affinché la procedura funzioni bene almeno per il 95% del tempo.
L'errore di L_C deve essere stimato simultaneamente con la posizione del *rover*, esso generalmente non può essere considerato in anticipo ben noto a sufficienza.
Altre incognite sono gli errori nella correzione troposferica e nelle effemeridi trasmesse. Come nella modalità in tempo reale, un filtro di Kalman viene utilizzato per ottenere una soluzione congiunta per la posizione attuale e tutte le altre incognite (in post-

[16] L'interpolazione è un procedimento matematico per la ricostruzione di una serie incompleta di dati. E' di pratico utilizzo con tabelle (di dati, funzioni trigonometriche, logaritmiche, attuariali etc.) quando si voglia calcolare un valore non riportato in esse. Il metodo dell'interpolazione lineare approssima i valori mancanti in un intervallo con una linea retta che unisce i due punti conosciuti agli estremi dell'intervallo: si può quindi calcolare un valore approssimato della funzione di un punto x qualsiasi tra due punti in cui la funzione sia nota commettendo un errore che è tanto più piccolo quanto l'intervallo è stretto.
[17] Quando un errore viene corretto, la correzione genera un altro errore.

processamento si dovrebbe andare oltre utilizzando arrotondamenti). I dati sono misure di fase in doppia frequenza e, quando possibile, *pseudorange* ottenuti da combinazioni *iono-free* di L1.
Questa soluzione congiunta è stata ottenuta utilizzando un particolare software per il posizionamento statico e cinematico preciso su lunghe distanze, modificato per applicare correzioni ionosferiche. Tale software è stato usato ripetutamente per calcolare posizioni statiche e cinematiche con precisione sub-decimetrica, per distanze maggiori di 100 km. Tutte le incognite richieste per il presente lavoro fanno già parte delle equazioni di osservazione. In particolare, le tecniche che trattano *range* lunghi che coinvolgono ambiguità "*float*", che utilizzano L_C come tipo di dato principale, risolvono questi errori come "numeri reali". Il calcolo fornisce anche la precisione per ogni errore L_C stimato, per decidere quando questa stima è precisa abbastanza da poter essere tranquillamente usata.

(b) Risolvere le ambiguità di L1 ed L2. Una volta che le ambiguità intere *widelane* N_w sono note, si può sfruttare il seguente vincolo corretto, l'osservabile ionosferica ambigua LI=L1-L2 deve soddisfare:

$$LI - N_1(\lambda_1 - \lambda_2) + N_w \lambda_2 - \text{corr. ion.} =$$
$$= \text{rumore}(L1) + \text{errore}(\text{corr. ion. per } LI)$$

Da cui segue che:

$$N_1 = \text{Intero più vicino}\left[\frac{(LI - N_w\lambda_2)}{\lambda_1 - \lambda_2}\right],$$

cosicché la somma del rumore e del errore di rifrazione ionosferica in LI è pari a meno della metà di $|\lambda_1 - \lambda_2|$, oppure è pari a 2.7 cm. Finalmente:

$$N_2 = N_1 - N_w$$

Una volta che N_1 e N_2 sono state trovate, l'esatto errore di L_C può essere calcolato ed assimilato nel filtro di Kalman, come (pseudo-)osservazione aggiuntiva. Finché il filtro in questione è del tipo di convarianza usuale, un po' di incertezza dovuta a rumore può essere assegnata ad ogni osservazione. In questo caso, σ = 1 cm è stato scelto come valore convenientemente piccolo.
La procedura descritta non richiede una ricerca degli interi purché le incertezze siano, entro certi limiti, piccole, in particolare l'effetto del rumore è stato ridotto mediando i dati. Il che è stato fatto facilmente, perché è una caratteristica della tecnica con *range* lunghi, che usa compressione (media) dei dati per economizzare tempo e spazio sul disco rigido. Tramite tecniche "*trial and error*", intervalli di due minuti per la media dei dati sono stati scelti. In generale, il tutto ha richiesto circa tre aggiornamenti consecutivi del filtro con misure di fase acquisite in media ogni due minuti, per rendere gli errori di L_C stimati e cedenti la precisione di 20 cm, richiesta per risolvere le *widelane*. (Per risoluzioni più veloci, una ricerca degli interi dovrebbe essere migliore, se teniamo propriamente conto delle varie incertezze.) Nessuna ambiguità è considerata risolta senza che il valore assoluto del contenuto LI (corretto, mediato, e presumibilmente esente da errori) sia meno di 1.5 cm. Nessun errore L_C risolto è stato accettato ed assimilato nel filtro di Kalman prima che esso abbia superato un test di ipotesi di errore nullo. E, mentre molti errori di L_C dovrebbero essere soddisfacentemente risolti ogni due minuti, solo quelli che non sono già noti con una precisione migliore di 3 cm sono stati aggiornati ai loro valori risolti

4.10.3 Ricerca OTF delle ambiguità GPS basata sulla Logica Sfumata[18] (Fuzzy-Logic)

La chiave per il posizionamento GPS OTF (On the Fly trad. al volo), per la navigazione ed il controllo dei veicoli con alta precisione (livello centimetrico) è la soluzione corretta del problema delle ambiguità. Questo significa che il set di interi corretto deve essere identificato da un numero a volte molto grande di candidati, entro un periodo molto breve. Numerosi algoritmi di ricerca sono stati scoperti ed introdotti da vari autori allo scopo di effettuare questa operazione.
Tali algoritmi funzionano bene entro certi limiti di validità che sono stati imposti per differenti ragioni:

- modelli matematici incompleti dovuti ad insufficiente conoscenza del modo reale;
- effetti sistematici incogniti nel materiale di osservazione;
- qualità sconosciuta delle osservazioni del processo stesso.

In breve: considerando tale problema vi sono delle incertezze che non vengono trattate propriamente a causa di informazioni imperfette.
La teoria degli insiemi e dei sistemi sfumati (*fuzzy*) è un approccio per trattare tali incertezze nei modelli matematici. In questo documento viene descritto un approccio che si basa sulla logica sfumata per identificare il corretto set di interi per risolvere le ambiguità presenti. Questo approccio si basa su un software ed un algoritmo OTF preesistenti, ed è stato sviluppato ed implementato in modo da migliorare l'identificazione del set di interi corretto. Lo scopo è di rendere la risoluzione OTF dell'ambiguità più affidabile e di estendere il campo di applicabilità del software. Una comparazione tra le statistiche originali e quelle ottenute con questo approccio è stata effettuata. Le differenti proprietà dei due approcci sono state discusse ed è stata inclusa la valutazione dei risultati dei dati reali.

Risoluzione integrata OTF dell'ambiguità

La tecnica IOTF (OTF di risoluzione integrata dell'ambiguità *Integrated* OTF) venne sviluppata nel '93 allo scopo d'effettuare posizionamenti molto precisi di un'antenna GPS in moto, persino a grandi distanze (da 50 km in su) dalle stazioni di riferimento. In aggiunta l'algoritmo può essere utilizzato in tempo reale.
I principali passi del processo IOTF entro una singola epoca d'osservazione sono riassunti di seguito in una tabella. La parte essenziale di questo metodo consiste in una sofisticata procedura di ricerca per identificare i valori corretti delle ambiguità *widelane* (2° passo). Da epoca ad epoca, solo i valori delle ambiguità fissate (e nessun'altra informazione) vengono trasferiti.

[18] Per un'introduzione alla teoria della logica sfumata si vada all'appendice C.

1. Soluzioni differenziali di codice

→ posizione approssimata X_C, C_{XC},definisce volumi di ricerca ellissoidici

2. Ricerca delle ambiguità widelane

- composizione di una lista di tutte le alternative: tutte le triplette di interi, quali ad esempio le potenziali ambiguità widelane, disposte dentro un volume di ricerca ellissoidico
- correzione sequenziale ai minimi quadrati:
  1. soluzione iniziale X_C, C_{XC}
  2. aggiornamento con il principale gruppo widelane di fase
  3. aggiornamento con il gruppo secondario widelane di fase

  → procedura di test di tutte le alternative tramite analisi dei residui

3. Posizionamento con le osservabili di fase widelane

4. Stima delle ambiguità di una 2ª frequenza

→ con le ambiguità widelane note, ad esempio $NL1 = f(\Phi_{ionofree}, N_{WL})$

5. Posizionamento con osservabili di fase iono-free

(o qualsiasi combinazione di fase L1/L2)

Tecnica IOTF: passi principali in una singola epoca

Delle due azioni critiche; 1ª ricerca delle ambiguità *widelane* e 2ª stima delle ambiguità di una seconda frequenza (e da questo soluzione completa del problema delle ambiguità a doppia frequenza), solo la parte *widelane* viene considerata in questo documento.
La ricerca delle ambiguità *widelane* significa la scelta di un set di interi come valori corretti da una lista di candidati, che si trovano tutti all'interno di un volume di ricerca ellissoidico. Una serie di test è stata applicata ad ogni set di interi candidato in modo da identificare quello corretto. Questo processo di selezione può essere considerato in termini di teoria della decisione un MADM (*Multi Attribute Decision Making* trad. Effettuazione di Decisioni sulla base di Attributi Multipli), si veda la tabella sottostante.
Nel caso della IOTF le alternative sono i set di interi candidati, mentre gli attributi sono costituiti dai vari test, che sono distribuiti secondo un ordine sequenziale in modo da analizzare i residui.

		Attributi 1 … j … n	Tasso totale (or)
Alternative	1 … i … m	***tassi locali*** (lr) (lr_{i1} … lr_{ij} … lr_{in}) ***aggregazione***	or_1 … or_i … or_m
			selezione della migliore alternativa

Come si può notare dalla precedente tabella tre passi principali costituiscono il MADM: i tassi locali, l'aggregazione, e la selezione. Tramite i tassi locali viene valutata la performance di una alternativa i rispetto un attributo j, cosi facendo viene costituito un vettore di tassi locali per ogni alternativa. Questo vettore sarà allora aggregato tramite un operatore adatto ad ottenere il tasso totale, grazie alla comparazione di tutti i tassi locali la selezione della migliore alternativa è facilitata. Ovviamente l'interazione propria tra il metodo dei tassi locali e un operatore di aggregazione è un punto cruciale per il successo del MADM.
Nel caso di ricerca dell'ambiguità *widelane*, solo una delle alternative è corretta ed accettabile, tutte le altre conducono a falsi risultati. Per questo motivo la selezione della migliore alternativa si riduce all'identificazione della sola alternativa corretta. Nel paragrafo successivo il metodo di ricerca originale della tecnica IOTF delle ambiguità *widelane*, è stato chiamato test *sharp*[19] o *crisp*[20], ed è stato discusso in termini di MADM.
A causa di inconvenienti essenziali, cioè ridotta affidabilità in condizioni di osservazione estreme, un approccio alternativo (in cui le decisioni del test vengono fuzzificate) è stato sviluppato.

Test crisp

L'ordine sequenziale delle epoche delle osservazioni *widelane* di fase è ottenuta per tutti i set di interi candidati. La rispettiva soluzione ed i residui di un singolo set sono soggetti ad otto verifiche e reiezioni (test VR). I criteri dei test sono basati su assunzioni statistiche e su soglie empiriche.

MADM crisp

Seguendo la terminologia MADM la ricerca *crisp* di ambiguità *widelane* viene effettuata secondo i passi così riassunti:

1. La decisione dei test VR, tramite tassi locali, viene compiuta secondo il modo *crisp*, il che significa che alla decisione viene attribuito un valore del set *crisp* {0,1}:

$$lr_{ij} = \begin{cases} 1 & \text{se il set di interi i supera il test j} \\ 0 & \text{else} \end{cases}$$

Così 1 vuol dire che un attributo è soddisfatto dall'alternativa, 0 significa attributo non soddisfatto.

2. L'aggregazione per raggiungere il tasso totale or_i per la i-esima alternativa viene ancora compiuta secondo il modo *crisp*:

$$or_i = Agg\left(\ldots lr_{ij} \ldots\right) = \begin{cases} 1 & \text{se tutti gli } lr_{ij} = 1 \\ 0 & \text{else} \end{cases}$$

[19] Intenso, forte.
[20] Deciso.

Questo significa, che per ogni set di interi candidato esso sarà verificato, solo se tutti i test sono stati superati.

3. Dopo che tutte le alternative sono state processate il numero n_{ac} dei set di interi accettati (aventi tasso totale pari ad 1) viene contato. Tre casi vengono distinti durante il passo di selezione:

n_{ac} = 0 nessun set di interi è stato accettato, le ambiguità *widelane* non possono essere determinate;
n_{ac} = 1 un set di interi è stato accettato e scelto come migliore alternativa, cioè esso è il valore corretto dell'ambiguità *widelane*;
n_{ac} > 1 più di una alternativa con il tasso totale massimo pari ad 1 è presente, non può essere effettuata una selezione;

Affinché la ricerca *crisp* delle ambiguità *widelane* identifichi un set di interi corretto, deve verificarsi che questo set soddisfi tutti i criteri del test ed in più dovrebbe essere il solo a farlo. La formulazione MADM è stata scelta per porre un analogia tra questo approccio e quello alternativo che verrà descritto più avanti.

Esempio

Per dare un idea del tasso di successo del *crisp* test sono state misurate simultaneamente quattro *baseline* (B1 … B4) di differente lunghezza. Dato che le coordinate di queste *baseline* sono note con precisione le ambiguità *widelane* corrette potrebbero essere computate e usate per verificare i risultati del metodo *crisp*. Sono stati considerati 5 casi, si veda la tabella 1 sottostante.
Allo stato F2 il metodo *crisp* ha pieno successo, mentre F1 ed F4 denotano che nessuna decisione può essere presa, in quanto nessuna significa successo o falsa soluzione. Sia F3 che F5, denotano che l'algoritmo sceglie set di interi sbagliati invece che corretti, i set sbagliati vengono sommati e forniscono il tasso di soluzione falsa. Le percentuali di questi casi rispetto a tutte le epoche si possono leggere nella tabella 2 successiva.

Tabella 1 - Verifica dei casi

caso	n_{ac}	verifica
• F1	n_{ac} = 0	nessuna soluzione selezionata
* F2	n_{ac} = 1	set di interi corretto selezionato
→ F3	n_{ac} = 1	falso set di interi selezionato
• F4	n_{ac} > 1	set di interi corretti e falsi accettati
→ F5	n_{ac} >1	solo set di interi falsi accettati

Tabella 2 - Percentuale di verifica dei casi: test crisp

baseline	lunghezza dei baseline	F1 [%]	F2 [%]	F3 + F5 [%]	F4 [%]
B1	19 km	9,4	73,1	5,5	18,6
B2	30 km	27,7	49,5	16,9	11,9
B3	38 km	20,8	60,3	7,8	16,4
B4	48 km	38,3	38,3	14,2	15,0

Intervallo temporale di tre ore con tasso di campionamento di 30 secondi ed angolo di elevazione di 15°;

La ricerca delle ambiguità *widelane* è stata effettuata per ogni singola epoca. E' riscontrabile una elevata variabilità nel tasso di successo e di errore da un *baseline* all'altro, cioè da un set di dati a un altro osservando la tabella 2. C'è più variabilità che ci si aspetti tendendo a diminuire il tasso di soluzione con *baseline* più lunghi. Nelle applicazioni reali, dove la posizione della antenna GPS mobile non è nota e varia da epoca ad epoca, una accurata verifica dei dati raccolti è assolutamente necessaria. L'insufficiente affidabilità delle soluzioni è ancora il problema più grande è più difficile da affrontare, nel posizionamento OTF su lunghe distanze.

Discussione

Il suddetto modello matematico IOTF è l'approssimazione ai minimi quadrati delle doppie differenze (DD) delle osservazioni GPS. Si assume che la tecnica DD elimini tutti gli errori sistematici, le osservazioni DD sono considerate variabili casuali con funzioni di distribuzione normali (cioè distribuite secondo la campana di Gauss). Se l'assunto è vero, ci si aspetta che i residui, ottenuti dal uso di un set di interi corretto per le ambiguità *widelane* (WL), varino intorno a zero come rumore casuale, e che la somma pesata dei quadrati dei residui sia minima (comparata con quelle ottenute usando set di falsi interi).
In realtà ci troviamo di fronte ad altri problemi, in quanto all'aumentare delle distanze e sotto severe condizioni atmosferiche e di *multipath*, le osservazioni DD sono affette da effetti sistematici significativi. Anche se in molti casi ne la grandezza ne il comportamento nel tempo di tali errori sistematici è noto, essi sono introdotti nel processo di arrotondamento. Cosi facendo la tecnica dei minimi quadrati produrrà un gran numero di effetti sistematici contraddicendo il modello statistico. Oltretutto i minimi quadrati hanno la proprietà di influenzare gli effetti sistematici di tutti i residui, questo può condurre alla reiezione del set di interi corretto nella procedura di test.
In più, se falsi interi di ambiguità WL sono introdotti nell'arrotondamento, l'effetto combinato dei valori sbagliati di ambiguità con gli errori sistematici nelle osservazioni di fase WL possono essere un'amplificazione o una diminuzione dell'ampiezza dei residui. Conseguentemente occorrono epoche, in cui set di interi falsi producono residui più piccoli di quelli prodotti con i set corretti e possono persino produrre la somma minima pesata dei quadrati dei residui. Durante tali epoche il rischio che il set di interi corretto venga rifiutato o vengano accettati set di interi falsi è molto elevato, il che conduce ai casi F1, F3, F4, ed F5 già descritti in precedenza.
Un altro problema può essere citato: facendo posizionamento cinematico OTF si dispone di piccole ridondanze, il che è in contraddizione con la parte statistica del modello. Considerando soltanto una stazione di riferimento, la ridondanza di una singola epoca sarà compresa tra 4 e 8-10 osservazioni (tutto dipende dall'angolo di elevazione, dalla visibilità e dalla costellazione satellitare). Dato che i criteri di alcuni test sono assenti funzioni statistiche di distribuzione, essi mostrano alcune debolezze.
Ora il problema sta nel valutare quanto il *crisp* test è affetto da tali incertezze. Il metodo *crisp* può essere considerato come un metodo assoluto nel senso che, solo il valore totale dei tassi, che può essere anche pari a zero, viene preso in considerazione. Non ci sono altre comparazioni possibili tra le alternative. Assegnando i valori {0,1} alle decisioni ottenute dai test, non abbiamo informazioni su come il criterio di un test è soddisfatto o su come un test fallito è contenuto in un tasso locale, e questo non viene trasferito al tasso totale.
Così facendo, il metodo di selezione si limita a contare le sole alternative il cui tasso totale è pari ad 1. Nessuna affermazione sull'affidabilità della selezione può essere fatta all'interno del *crisp* MADM stesso. Questo è un problema critico in quanto gli effetti

sistematici che abbiamo discusso portano all'aumento della grandezza di disturbi più violenti nel processo di decisione. Una decisione senza un'affermazione propria sulla sua affidabilità avrà poco valore.
Un altro inconveniente del metodo *crisp*, sta nel fatto che tutti i test hanno lo stesso peso nell'aggregazione, che finisce con una sorta di veto: se un test, persino di poca importanza, fallisce, la possibilità che l'alternativa sia accettata verrà esclusa. Dal punto di vista esperto questa è una brutta situazione, in quanto non è possibile evitare questo caso estendendo il criterio del test, dato che questo consentirebbe a molti set di falsi interi di essere accettati. Ciononostante un graduale "rating" locale sembra aprire la strada ad un più sofisticato giudizio delle differenti alternative.

Test Fuzzy

Il motivo per applicare il metodo basato sul concetto dei set *fuzzy* è basato sull'analisi degli inconvenienti del metodo *crisp*. Le aspettative e i differenti aiuti concernenti l'implementazione di un metodo *fuzzy* di decisione sono:

- ottenere una graduale valutazione della singola alternativa, che riflette la plausibilità che essa sia il set di interi corretto;
- aumentare l'affidabilità delle decisioni tramite una ispezione più dettagliata della valutazione delle alternative;
- raggiungere un metodo di ricerca delle ambiguità WL più robusto rispetto alla differente qualità dei dati (differenti condizioni atmosferiche, lunghezza dei *baseline* etc.);

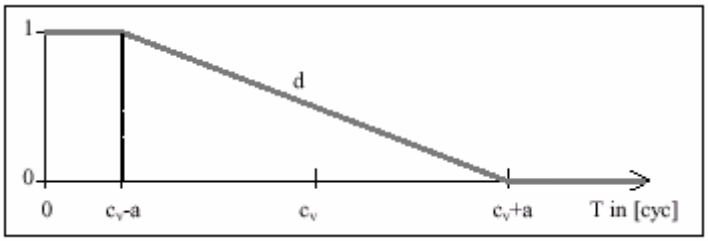
MADM Fuzzy

Allo scopo di ottenere maggiori informazioni dai singoli test, una fuzzificazione delle decisioni dei test è stata effettuata. Invece della sola assegnazione dei valori {0,1} circa il fatto che la decisione abbia superato o no il test, viene posta una funzione membro per ogni test allo scopo di descrivere gradualmente la validità di una alternativa per essere considerata adatta alle aspettative, in merito al comportamento del set di interi corretto.
Questo test è stato effettuato comparando la norma di L1 dei residui delle osservazioni di fase con una soglia empirica, o più esattamente: assegnando a T, il significato aritmetico dei valori assoluti di tutti gli n residui, espressi in unità di ciclo WL, che superano la soglia c_v:

$$T = \frac{1}{\lambda_{WL}} \frac{1}{n} \sum_{i=1}^{n} |v_i| \leq c_v$$

Per fuzzificare la decisione "d" del test, una funzione membro viene posta come funzione rampa con un parametro "a", nel seguente modo:

$$d = \begin{cases} 1 & \text{se } T \leq c_v - a \\ 1 - \dfrac{T - (c_v - a)}{2a} & \text{se } c_v - a \leq T \leq c_v + a \\ 0 & \text{se } T \geq c_v + a \end{cases}$$



Funzione membro di un test decisionale fuzzificato

Come si può notare, solo il test decisionale è fuzzificato, l'idea dei test rimane invariata così come rimangono invariate le già citate assunzioni (uguali se di natura statistica o empirica).
L'aggregazione MADM Fuzzy segue una strategia che pesa i tassi locali in accordo con le valutazioni di un esperto su una coppia di attributi di importanza relativa. Il MADM Fuzzy è composto dai seguenti passi:

1. Da sette ad otto test VR del metodo *crisp* per ottenere il tasso locale usando la rispettiva funzione membro per assegnare gradualmente al test un valore dell'intervallo [0,1]. Un tasso locale come questo descrive se l'alternativa soddisfa completamente (1) o per nulla (0) il criterio del test, o se in qualche modo si trova a metà strada tra queste due alternative. Il test numero otto (tasso dell'alternativa della somma minima pesata dei quadrati dei residui) non è usato per il tasso totale dell'alternativa ma come parte del metodo di selezione. Si veda il terzo passo.
2. Usando i pesi w_j dei test, ottenuti con il metodo di Saaty, che descrive l'importanza dei test relativamente ad ogni altro, i tassi locali pesati sono scritti come:

$$\hat{l}r_{ij}^{\,w_j}$$

   i pesi sono normalizzati con $\Sigma w_j = 1$, questo w_j proviene da [0,1]. Per cui i tassi locali provenienti da test di minore importanza saranno esponenzialmente più forti di quelli più importanti.
   Tutto ciò ha senso dato che i tassi locali pesati saranno aggregati tramite l'operatore minimo:

$$or_i = Agg\left(\ldots\hat{l}r_{ij}\ldots\right) = \min\left(\ldots\hat{l}r_{ij}\ldots\right)$$

   questo significa, che il più debole tasso pesato locale include il tasso totale per l'alternativa.
3. Per il processo di selezione, al primo valore massimo di tutti i tassi totali, or_{max}, e la somma minima pesata dei quadrati dei residui, Ω_{min}, di tutti i possibili set di interi determinati. or_{max} deve avere estenzione minima, altrimenti nessuna scelta sarà fatta all'epoca attuale, in quanto questo indica che nessun set d'interi include un tasso totale accettabile. Questo può accadere in epoche con effetti sistematici estremamente grandi.

Se or_{max} è accettabile, tutti i possibili "i" sono accettati, se essi soddisfano le seguenti condizioni:

$$\text{(I)}\ T_{\mathrm{or,i}} = \frac{or_i}{or_{\max}} \geq soglia_{or}$$

$$\text{(II)}\ T_{\Omega,\mathrm{i}} = \frac{\Omega_i}{\Omega_{\max}} \leq soglia_{\Omega}$$

$$\text{(III)}\ T_{\mathrm{cross}} = \frac{T_{\Omega,i}}{T_{or,i}} \leq soglia_{cross}$$

$$\text{con}\ T_{\mathrm{or,i}} \in [0,1], T_{\Omega,\mathrm{i}} \in [1,\infty), T_{\mathrm{cross}} \in [1,\infty)$$

La condizione (I) assicura che solo i candidati considerati, i cui tassi totali sono sufficientemente vicini a or_{max}, mentre la condizione (II) è una verifica della quantità più importante, la somma pesata dei quadrati dei residui, che non è lontana da quella minima. La condizione (III) è introdotta per rifiutare il set di interi che passa le soglie delle condizioni (I) e (II) solo molto debolmente. Tutte le soglie sono determinate empiricamente, i valori usati per le valutazioni sono elencati nella tabella successiva. Il numero n_{ac} delle alternative, che soddisfano tutte e tre le condizioni, viene contato, e tre casi vengono distinti, similmente al metodo crisp:

$n_{ac} = 0$ nessun set di interi è stato accettato, le ambiguità widelane non possono essere determinate;
$n_{ac} = 1$ un set di interi è stato accettato e scelto come migliore alternativa, cioè esso è il valore corretto dell'ambiguità widelane;
$n_{ac} > 1$ più di una alternativa viene accettata, nessun singolo set di interi è chiaramente superiore, non può essere effettuata una scelta;

Soglie nel passo di selezione del MADM Fuzzy

Soglia	Valore
Minimo per or_{max}	0,5
$Soglia_{or}$, condizione (I)	0,8
$Soglia_{\Omega}$, condizione (II)	2,0
$Soglia_{cross}$, condizione (III)	1,7

Il metodo di aggregazione descritto nel secondo passo del MADM Fuzzy richiede, che le rampe delle funzioni membro dei test di lunga estensione, togliendo solo una piccola zona per un tasso locale pari ad uno, ed estendendo in avanti l'angolo di reiezione, che un set di interi corretti non troverà mai nella zona membro di grado zero. Tutto ciò serve ad ottenere una gradazione adatta dei tassi locali e ad evitare, che per un set di interi corretto un singolo tasso locale pari a zero si verifichi. Altrimenti questo dovrà includere un tasso totale pari a zero, dovuto al tipo di operatore di aggregazione (operatore minimo) scelto.
Per l'aggregazione vi sono numerosi metodi applicabili, è chiaro che la complessa interazione tra la fuzzificazione, il loro peso e il metodo di applicazione è decisivo per il successo della ricerca *fuzzy* delle ambiguità.

Esempio

Il set di dati considerato nel precedente esempio era valutato con la procedura MADM Fuzzy. Dovuta al fatto, che il processo di selezione *fuzzy* distingue tre casi $n_{ac} = 0$, $n_{ac} = 1$ e $n_{ac} > 1$, come si verifica anche per la selezione *crisp*, gli stessi cinque casi della prima tabella sono verificati. Le percentuali risultanti rispetto a tutte le epoche valutate sono elencate nella tabella successiva. La variabilità da un set di dati all'altro ed anche la tendenza a decrescere del tasso di soluzione all'aumentare della differenza è simile ai risultati del test *crisp*. Invece le percentuali dei cinque casi sono differenti, poche epoche senza set di interi accettati (caso F1), mentre il tasso di successo (caso F2) è significativamente aumentato. Ma qualcosa aumenta anche il tasso di errore (casi F3 ed F5).

Percentuale di epoche senza set di interi accettati.

Tabella -Percentuale di verifica dei casi: test crisp

baseline	lunghezza dei baseline	F1 [%]	F2 [%]	F3 + F5 [%]	F4 [%]
B1	19 km	1,1	81,7	3,2	13,9
B2	30 km	5,8	61,8	19,1	13,3
B3	38 km	3,1	73,9	8,6	14,4
B4	48 km	5,8	56,1	19,7	18,3

Intervallo temporale di tre ore con tasso di campionamento di 30 secondi ed angolo di elevazione di 15°.

Fine – Deo gratias semper