

Analisi statistica delle serie temporali di dati provenienti dalla rete di stazioni permanenti dell'UNIPA

G. Dardanelli, P. Orlando, M. Sciortino

DIRAP – Dipartimento di Rappresentazione, Università di Palermo
Viale delle Scienze, 90118 Palermo, tel. 09123896228 fax 091588853
email: gino.dardanelli@unipa.it, pietroorlando@unipa.it; marco.sciortino@alice.it

Riassunto

In questo lavoro ci si è proposti di verificare la qualità dei dati provenienti dalle stazioni permanenti appartenenti alla rete dell'Università di Palermo, creata nell'ambito del progetto PRIN2005 dal titolo: "Reti di stazioni permanenti GPS per il rilievo in tempo reale in impieghi di controllo e emergenza" e gestita dal Dipartimento di Rappresentazione. Attualmente sono disponibili serie temporali di soluzioni giornaliere. Queste permettono di individuare fenomeni ed eventi di varia natura come quella geofisica, geologica, meteorologica e, proprio in virtù della numerosità di questi dati si possono rilevare fenomeni di rumore più o meno complessi che possono essere apparentemente di natura deterministica e che senza l'uso di opportune tecniche di analisi possono essere scambiati per tali.

I tipi di rumore che comunemente caratterizzano i dati di stazioni permanenti possono essere White noise, Random Walk noise, Flicker noise e Auto regressive noise. L'analisi statistica è finalizzata all'individuazione del tipo di rumore, alla stima dei parametri del modello di interpolazione, alla valutazione della sua correttezza, al suo aggiornamento in caso di individuazione di errori grossolani (outlier) o sistematici e alla stima dei parametri del modello stocastico adottato. La metodologia seguita per la caratterizzazione stocastica basata sull'analisi delle serie temporali di coordinate intende essere un nuovo approccio facilmente implementabile su MATLAB.

Abstract

Today GNSS coordinate time series analysis was evaluated to determine accurate position. The aims of this project is a preliminary combination of raw data to obtain the initial accurate parameter estimates of the stations in order to compute error bounds for all parameters of the model. Case of study is *UNIPA NRTK GNSS*: this network is located in Italy, central western of Sicily and is composed of nine permanent stations.

The coordinates of the reference stations were established in *IGS05* (epoch 2005.0) and *ETRF89* (epoch 1989.0) system. The pre-processing strategy used provided of the coordinate time series in which outliers and discontinuities were identified. After outlining the GPS coordinate time series analysis strategy this paper shows initial results for coordinate time series for one year period.

Statistical analysis aimed at identifying the type of noise, the estimated parameters of the model interpolation, the evaluation of its correctness, its update in the event of detection of gross errors (outliers) or systematic and to estimate parameters stochastic model adopted. The methodology for the stochastic characterization based on analysis of time series of coordinates will be a new approach easily implementable on MATLAB.

Introduzione

Da diversi anni, i ricercatori del Dipartimento di Rappresentazione (DIRAP) dell'Università di Palermo hanno la gestione della rete di stazioni permanenti GNSS della Sicilia Centro Occidentale. Questa rete è organizzata con nove stazioni localizzate rispettivamente ad Agrigento, Alcamo (TP), Caltanissetta, Campobello di Mazara (TP), Palermo, Partinico (PA), Prizzi (PA), Termini Imerese

(PA) e Trapani ed è coordinata dal centro di controllo sito presso la Facoltà di Ingegneria dell'Università di Palermo. La rete fornisce, in via sperimentale, dati per servizi di posizionamento sia in tempo reale che in post-elaborazione (Dardanelli et al., 2008) attraverso un Webserver, di imminente pubblicazione; le stazioni di Campobello di Mazara (TP) e di Termini Imerese (PA) sono già state inserite nella RDN gestita dall'IGM, per la definizione del nuovo *datum* geodetico italiano, mentre la stazione di Palermo è di imminente inserimento.

Partendo da queste considerazioni, in particolare, l'analisi di serie temporali di coordinate provenienti da stazioni GNSS risulta un utile strumento da applicarsi, con diverse finalità, sia a monte che a valle del processo di trattamento dei dati misurati.

Come è noto in letteratura, le serie temporali storiche risentono spesso di imperfezioni quali dati mancanti, outliers e discontinuità nei valori medi; l'argomento è di interesse sia per la frequenza in cui questi problemi si riscontrano nelle serie, sia per le ripercussioni sugli esiti delle analisi e delle applicazioni condotte con i dati affetti da errore.

Le discontinuità nel valore medio, possono essere effetto di *offset* strumentali e delle loro correzioni, di modifiche nella stazione di rilevamento (posizione, tipologia degli strumenti) o anche di variazioni dell'ambiente che circonda la stazione; nella serie queste discontinuità si presentano come gradini o trend. Come per gli *outliers*, la registrazione indica un valore differente dal valore reale della grandezza misurata ma, a differenza di questi, nelle discontinuità della media l'errore è sistematico e modifica i dati di una quantità che persiste per un certo periodo, mantenendo approssimativamente stabili le altre caratteristiche dei dati, quali ad esempio la differenza di passo uno e parte degli eventuali andamenti stagionali. Gli spostamenti nel valore medio possono affliggere le serie di dati per lunghi periodi, anche diversi anni, come ad esempio nel caso di cambiamenti della posizione o della strumentazione della stazione, o di mutamenti dell'ambiente circostante.

Procedura di analisi delle serie temporali

Il periodo preso in considerazione in questo studio, tiene conto del fatto che tra la fine del 2009 ed i primi mesi del 2010 è iniziato il ciclo solare numero 24; si è scelto di utilizzare per la sperimentazione il periodo compreso tra il 28 Luglio 2008 ed il 28 Luglio 2009, in corrispondenza della fase finale del ciclo solare durante la quale l'entità dell'attività è risultata ai valori minimi.

La prima fase della sperimentazione ha riguardato il trattamento preliminare dei dati delle stazioni permanenti acquisiti con tre software, due dei quali sono correntemente utilizzati in ambito internazionale nel pretrattamento dei dati provenienti da stazioni GNSS. La SP di Alcamo non è stata presa in considerazione, poiché è stata solo recentemente aggregata alla rete e quindi non si dispone ancora di un range adeguato di dati.

I software utilizzati sono *TEQC* dell'UNAVCO, *QC2SKY* del Politecnico di Torino e *Network Deformation Analysis* (NDA), nella versione *Professional*, della Galileian Plus s.r.l.

Tralasciando i risultati dei primi due software, è interessante notare in questo lavoro le elaborazioni sviluppate con NDA, per la stima preliminare delle coordinate delle stazioni permanenti della rete *UNIPA*; questo software (Cersich et al., 2002) è stato sviluppato dalla Galileian Plus s.r.l., in collaborazione con il Politecnico di Milano e con l'Agenzia Spaziale Italiana (ASI).

Il processamento dei dati è stato impostato sia con soluzione su singola linea di base che con soluzione di rete, correzione della posizione del centro di fase dell'antenna, effemeridi precise dei satelliti GPS, oltre alle effemeridi relative al Sole e alla Luna ottenute dal *Jet Propulsion Laboratory* (JPL/EPH). Successivamente, sono state inserite le coordinate iniziali delle stazioni della rete, calcolate con Bernese 5.0 e quelle della rete IGS (Cagliari, Noto, Matera). Le stazioni sono poi state suddivise in due categorie: *free network point* e *constrained network point*. Con lo scopo di garantire maggiore precisione nella stima delle coordinate, sono stati caricati all'interno del programma i parametri relativi alle maree oceaniche. Successivamente si è scelto quale modellazione utilizzare per il calcolo e la correzione dell'errore ionosferico e dell'errore troposferico. Per modellare l'errore ionosferico è stato scelto il modello di *Klobuchar* (1996), utilizzando i parametri giornalieri forniti dal CODE dell'Osservatorio Astronomico dell'Università

di Berna, mentre per ciò che riguarda la modellazione della troposfera sono stati utilizzati sia il modello di *Saastamoinen* sia quello di *Hopfield*, entrambi modificati con la mapping function di *Niell*. Infine, per ciò che riguarda il fissaggio delle ambiguità si è scelto di utilizzare il metodo LAMBDA associato ad un test secondario sul ratio, che ad oggi risulta una delle metodologie maggiormente utilizzate in letteratura.

Analisi dei risultati

Per quanto riguarda i risultati derivanti dall'elaborazione dei dati, si farà riferimento ai valori ottenuti nel modello troposferico di *Saastamoinen*.

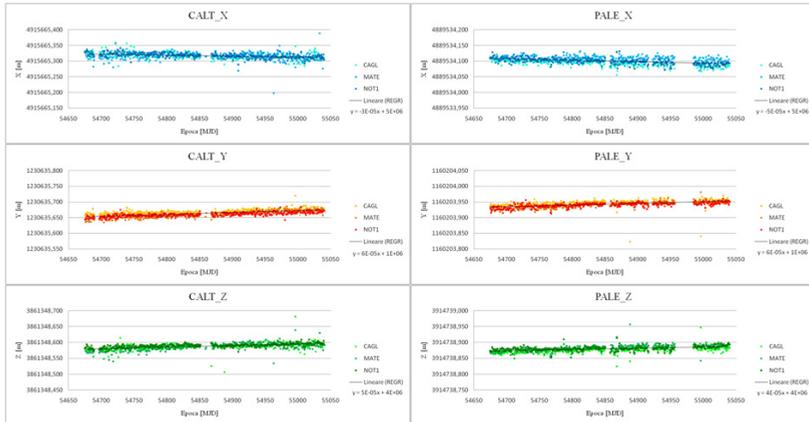


Figura 1 – Risultati per le stazioni CALT e PALE.

Nelle figure che seguono sono riportati gli andamenti preliminari delle serie temporali delle stazioni, considerando i valori determinati dalle tre stazioni di riferimento della rete IGS (Cagliari, Noto, Matera) nelle tre componenti X, Y, Z.



Figura 2 - Risultati per le stazioni PRIZ e TRAP

I migliori risultati si ottengono dalla distribuzione dei dati delle stazioni permanenti di Caltanissetta e Palermo che presentano delle serie temporali aventi andamento lineare pressappoco orizzontale, il che è indicativo a livello geodetico del fatto che i siti di installazione non hanno subito, durante l'arco temporale, deformazioni; a livello puramente tecnico ciò indica una presenza molto bassa di

rumore nelle osservazioni, che altrimenti tenderebbero a disporsi con andamento certamente più casuale (Figura 1).

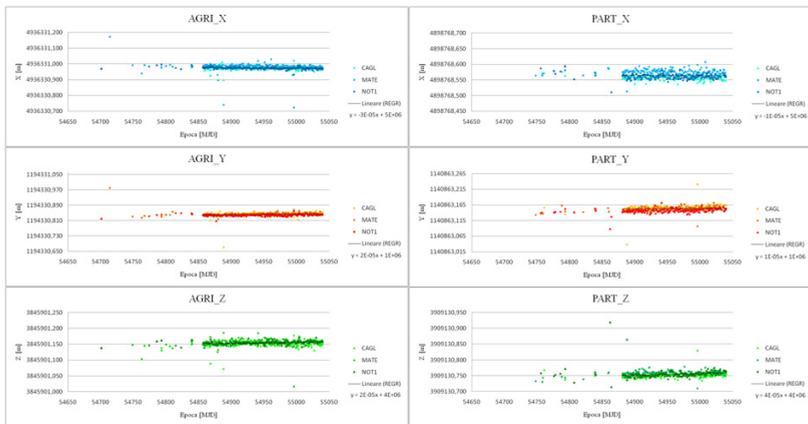


Figura 3 - Risultati per le stazioni AGRI e PART

Le stesse considerazioni non possono però certamente essere fatte per le stazioni permanenti di Prizzi e Trapani, che presentano delle serie temporali nelle quali i risultati si dispongono in maniera molto più casuale; a causa di ciò non è stato possibile rilevare un chiaro andamento, come nei casi precedenti (Figura 2).

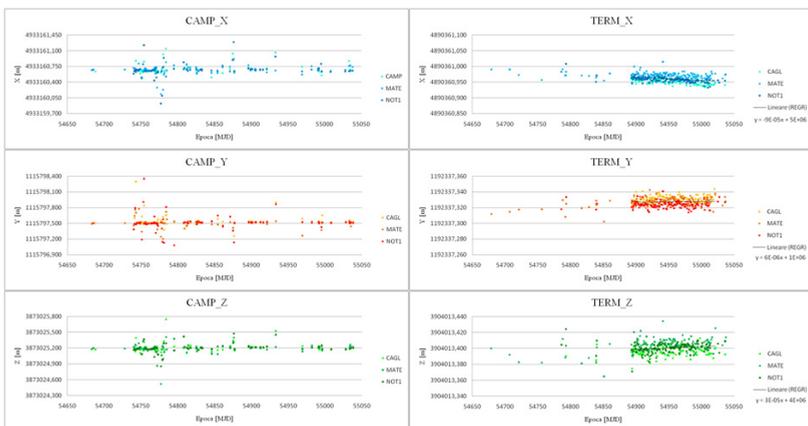


Figura 4 - Risultati per le stazioni CAMP e TERM (Saastamoinen)

In merito alle altre quattro stazioni, si può osservare che i risultati, si situano in posizione mediana rispetto ai casi estremi, di cui si è discusso in precedenza. Per le stazioni di Agrigento e Partinico si è ricavato un andamento lineare anche se limitato ad un arco temporale pari all'incirca alla metà di quello temporale analizzato (Figura 3).

Si ritrova un andamento simile a quello delle stazioni di cui sono stati esposti i risultati in precedenza anche per il sito di Termini Imerese, che appartiene alla RDN, malgrado in questo caso il numero di stime ottenute dal calcolo sia lievemente inferiore.

Dai dati provenienti dalla stazione di Campobello di Mazara, anch'essa facente parte della RDN, si sono ricavate invece delle serie temporali nelle quali non è stato possibile definire un andamento definito (Figura 4).

Approccio metodologico (analisi statistica)

Il problema fondamentale dell'analisi statistica delle serie temporali di stazioni permanenti GPS è quello di valutare le anomalie e i salti all'interno della serie riuscendo a capire quali tra queste anomalie sono legate all'andamento nel tempo della posizione della stazione GPS, come accade nel caso dei terremoti silenziosi (Miller et al., 2002), e quali invece dipendono da errori strumentali o da modifiche meteorologiche o climatologiche. Per poter isolare i salti del valor medio prodotti dai soli spostamenti, è necessario innanzitutto individuare, modellare ed eliminare i disturbi di diversa origine.

In questo lavoro verrà presentato un possibile approccio che consentirà di individuare e correggere le perturbazioni del valor medio. Eliminati questi disturbi, si analizzeranno gli spettri di potenza delle serie temporali della posizione di queste stazioni con lo scopo di determinare le caratteristiche del rumore presente in esse.

La mancata correzione di questi salti, oltre che rendere molto difficile la ricerca di possibili perturbazioni di origine sismogenetica, può rappresentare una fonte di errore non trascurabile nella stima della velocità tettonica con cui si propaga nel tempo la posizione della stazione.

E' necessario fare alcune riflessioni sulle possibili discontinuità: se l'entità dello scostamento dei valori medi è estremamente elevata risulta semplice identificare l'anomalia sia per via statistica che per via grafica. Quando lo scostamento è molto piccolo non è possibile identificarlo ma allo stesso tempo risulta modesto il difetto introdotto nella serie. Il problema è più complesso quando ci si trova in una situazione intermedia: lo spostamento non è tanto forte da essere palesemente una deviazione artificiale, ma è tale da produrre ripercussioni nell'uso dei dati. In questo caso occorre definire statisticamente, sulla base delle caratteristiche della serie la probabilità che esistano uno o più punti di discontinuità.

In statistica negli ultimi anni, sono stati sviluppati diversi metodi per l'identificazione delle anomalie artificiali nella media che prendono il nome di "metodi di omogeneizzazione". La quasi totalità dei metodi proposti sono studiati per verificare l'esistenza di un singolo punto di discontinuità in una serie. In genere quest'ultimo viene identificato analizzando i valori di una statistica riferita, direttamente o indirettamente, alla variazione della media dei dati tra la porzione di serie che precede il presunto punto di discontinuità e quella che lo segue. Ultimamente sono stati sviluppati nuovi metodi per l'individuazione di salti multipli all'interno di serie temporali.

Tra i metodi di ricerca di in presenza di salti multipli i più robusti sono il test SNH, il test MLR, e il metodo 2PR. Per questi metodi il procedimento per successive suddivisioni della serie produce buoni risultati quando i salti multipli sono strettamente gerarchici. Questo accade quando la serie (o la porzione di serie) esaminata contiene più discontinuità, ognuna delle quali determina una suddivisione della serie in due porzioni con medie diverse tra loro, ma vi è sempre una discontinuità per cui la differenza tra le due medie è significativamente maggiore di tutte le altre. Un esempio di discontinuità gerarchiche è costituito da un insieme di salti solo crescenti (o decrescenti); anche combinazioni di salti crescenti e decrescenti risultano spesso gerarchiche. Tipicamente non gerarchica è però una tipologia di discontinuità molto frequente, che si verifica ogniqualvolta un'anomalia nel funzionamento dello strumento di misura persiste per un certo intervallo di tempo prima di venire corretta; l'anomalia produce una coppia di salti complementari, che determina un intervallo di valori con media anomala, preceduto e seguito da intervalli con medie uniformi.

Esistono metodi appositamente ideati per trattare discontinuità multiple ma presentano, in alcuni casi, i medesimi difetti dei metodi per salto singolo, in altri danno seri problemi nell'identificazione del numero di discontinuità da considerare come significative (Menne e Williams Jr., 2005).

In questo articolo si presenta una metodologia per l'individuazione di intervalli di valori con media anomala, sviluppato nel 2009 per applicazioni a serie temporali caratterizzate da salti multipli. Il metodo opera variando la posizione di due punti di discontinuità, anziché di uno solo come nel metodo SNH, e per questo è stato chiamato con l'acronimo SNH2. L'individuazione dei salti è il

primo passo verso l'eliminazione degli outliers e la correzione dei dati della serie temporale al fine di procedere successivamente all'analisi spettrale.

Conclusioni

Si è visto che i risultati ricavati dal pretrattamento dei dati attraverso l'uso di tre differenti software scientifici TEQC, QC2SKY e NDA Lite abbia restituito piena corrispondenza dei risultati. In particolare, si sottolinea come in tal senso tutte le stazioni appartenenti alla rete abbiano presentato degli elevati livelli di qualità sia in termini di potenza del segnale ricevuto (SNR), che in termini di multipath e di numero di cycle slips.

La rete di stazioni permanenti dell'Università di Palermo anche se nata per scopi prettamente sperimentali presenta una qualità dei dati elevata assolutamente paragonabile a quella di altre reti operative sul territorio nazionale ed internazionale per scopi commerciali e/o scientifici.

La prima conclusione è che, relativamente alle condizioni imposte e a i dati utilizzati, i risultati sono indipendenti dal modello troposferico utilizzato. Si è dimostrato che, a parità di modello ionosferico, la scelta del modello troposferico di *Saastamoinen* e la scelta del modello di *Hopfield*, risulta essere ininfluenza sia sui valori delle stime delle coordinate delle stazioni permanenti che sui valori delle precisioni delle stesse.

Lo stesso non può dirsi per ciò che riguarda la dipendenza dei risultati dalla scelta del modello di correzione dell'errore ionosferico. Le analisi successive sono state dunque condotte attraverso l'uso del modello di *Klobuchar*, con i coefficienti del *CODE*. Si è visto come a seguito di tale scelta la precisione delle coordinate calcolate, valutata attraverso l'analisi degli scarti quadratici medi, sia aumentata in maniera assolutamente evidente.

Attualmente il Dipartimento sta portando avanti l'analisi statistica delle serie temporali secondo la metodologia esposta al fine di valutare e depurare i risultati da errori non correlati allo spostamento delle placche tettoniche.

Bibliografia

- Dardanelli G., Franco V., Lo Brutto M. (2008). La rete GNSS per il posizionamento in tempo reale dell'Università di Palermo: progetto, realizzazione e primi risultati, in: *Bollettino SIFET* (Società Italiana di Fotogrammetria e Topografia) n. 3-2008, pp. 107-124
- Shumway, Robert H., Stoffer, David S. (2006): *Time Series Analysis and Its Applications With R Examples*, Springer Editor, 2nd ed., XIII, 575 pp., Hardcover ISBN: 978-0-387-29317-2
- Chersich M., De Giovanni A., Osmo M. (2002): "NDA: un tool italiano per il processamento automatico di dati da reti GPS permanenti". In VI Convegno Nazionale ASITA, Perugia, Italy
- Miller M., Melbourne T., Johnson D.J., Sumner W.Q. (2002): *Periodic Slow Earthquakes from the Cascadia Subduction Zone*, Department of Geological Sciences, Central Washington University, Ellensburg, WA 98926, USA.
- Klobuchar J. A. (1996): "Global Positioning System: Theory and Applications". Volume I, Cap. XII: Ionospheric effects on GPS, pp. 485-515, American Institute of Aeronautics and Astronautics Inc.
- Niell A. E. (1996): "Global mupping functions for the atmosphere delay at radio wavelengths". *Journal of Geophysical Research*, Vol. 101, n. B2, pp. 3227-3246, Febbraio 1996