

Analisi delle componenti indipendenti (ICA) nello studio delle deformazioni del suolo in aree vulcaniche

Cataldo GODANO¹, Folco PINGUE², Milena BOTTIGLIERI¹, Mariarosaria FALANGA³,
Umberto TAMMARO², Prospero DE MARTINO², Francesco OBRIZZO²

¹ Dipartimento di Scienze Ambientali, Seconda Università di Napoli, via Vivaldi 43, 81100 Caserta, Italia. E-mail cataldo.godano@unina2.it

² Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli - Osservatorio Vesuviano, via Diocleziano 328, 80124 Napoli, Italia. E-mail pingue@ov.ingv.it

³ Dipartimento di Fisica, Università di Salerno, via S. Allende, 84081 Baronissi (SA), Italia. E-mail rosfal@sa.infn.it

Riassunto

L'analisi delle componenti indipendenti (ICA) è una tecnica recente e ben nota usata per separare miscele di segnali. Questa tecnica è stata applicata alle serie temporali delle deformazioni del suolo registrate alla rete permanente GPS dell'INGV-Osservatorio Vesuviano allo scopo di caratterizzare il livello di background delle deformazioni nell'area vulcanica napoletana. L'analisi ha rilevato la presenza di 5 segnali periodici indipendenti comuni a tutte le stazioni GPS, alcuni dei quali sono interpretabili come effetti mareali. I segnali residui ad ogni stazione rappresentano la deformazione del suolo locale. Sfortunatamente l'ICA non consente la valutazione delle ampiezze assolute delle componenti; pertanto non siamo in grado di ottenere le ampiezze residue ad ogni stazione. Quindi abbiamo usato un'analisi di stazionarietà allo scopo di evidenziare la presenza di eventuali deformazioni locali transienti. La tecnica ICA combinata con l'analisi di stazionarietà ha mostrato di essere uno strumento estremamente interessante per individuare trend di deformazione del suolo e potrebbe diventare molto utile nel monitoraggio delle aree vulcaniche.

Abstract

Independent component analysis (ICA) is a recent and well-known technique used to separate mixtures of signals. This technique has been applied to the ground deformation time-series recorded at the permanent GPS network of the Osservatorio Vesuviano-INGV in order to characterize the deformation background level in the Neapolitan volcanic area. The analysis revealed the presence of five independent periodic signals common at all the GPS stations; some of them are interpreted as effects of earth tides. The residual signal at each station represent the local ground deformation. Unfortunately the ICA cannot provide the absolute amplitude of the components, indeed we are not able to obtain a residual amplitude at each station. Then we used a stationarity analysis in order to investigate the eventual presence of local transient deformations. The ICA technique combined with the stationarity analysis has shown to be a very interesting tool for individuating ground deformation trends and could be very useful in volcanic areas monitoring.

Introduzione

Il problema della definizione del livello di background delle deformazioni del suolo è stato affrontato da molti autori e rappresenta una questione cruciale per la sorveglianza dei vulcani attivi. La sua stima consente, infatti, la definizione del livello di allerta per uno dei parametri della sorveglianza.

L'importanza dei movimenti del suolo in aree vulcaniche è dovuta all'osservazione che l'attività vulcanica, a causa sia di processi tettonici sia di movimenti di masse magmatiche, produce movimenti significativi della crosta superiore. Le tecniche geodetiche usate in geofisica,

determinano alcuni elementi peculiari di una serie di capisaldi nell'area studiata. I trend temporali di tali elementi forniscono informazioni utili alla comprensione della geometria e del meccanismo della sorgente delle deformazioni. Tra le metodologie usate per la sorveglianza delle deformazioni superficiali, i GPS hanno evidenziato enormi potenzialità nel monitoraggio dell'attività vulcanica. Questa tecnica ha un'accuratezza relativa dell'ordine di $10^{-6} - 10^{-7}$ (Silver et al., 1999; Czajewski, 2004). Molti autori hanno ottenuto risultati interessanti nell'analisi dei dati di deformazione registrati allo stesso sito in periodi differenti (vedi ad esempio Fernandez et al., 2004; Bonforte et al., 2004). Nondimeno è noto che i processi vulcanici possono modificare il campo di deformazione molto rapidamente rendendo inefficaci le tecniche di confronto tra misure discrete effettuate in tempi successivi. Cosicché, negli ultimi anni, traendo vantaggio dagli sviluppi tecnologici degli apparati GPS, l'uso dei sistemi di monitoraggio in continuo tramite GPS si è diffuso enormemente in tutto il mondo (si veda ad esempio Lowry et al., 2001; Pingue et al., 2003; Bonforte and Puglisi, 2006). È divenuta così di estrema importanza l'analisi delle sorgenti di rumore presente nelle serie temporali GPS in modo da isolare i fattori temporali propri delle sorgenti di deformazione (si veda tra gli altri Savage, 1988).

In questa nota noi adotteremo l'analisi delle componenti indipendenti (ICA) come una tecnica molto potente nel separare misture di segnali. In tal senso le deformazioni del suolo registrate sono state considerate come la sovrapposizione di un segnale di background, considerato comune a tutte le stazioni della rete GPS, a cui si aggiunge ogni deformazione di carattere locale. Qui noi intendiamo per deformazione di background una sorta di livello di deformazione stazionaria al di sopra della quale ogni transiente può essere considerato come indicativo di "qualcosa che sta cambiando nel sistema". Ovviamente la deformazione di background potrebbe essere affetta da errori sistematici che, comunque, possono essere risolti in uno stadio di pre-analisi. Infine vogliamo far notare che le deformazioni locali non possono essere separate tramite l'ICA (si veda la sezione successiva) e pertanto noi assumeremo l'esistenza di un segnale stazionario che sarà evidenziato dalla nostra tecnica ed eseguiremo un'analisi di stazionarietà per evidenziare l'eventuale presenza di transienti di deformazione locale.

Analisi delle componenti indipendenti

L'analisi delle componenti indipendenti (ICA) è un metodo per evidenziare fattori soggiacenti, o componenti, da dati multivariati (Hyvarinen et al., 2001). Essa è strettamente correlata al metodo chiamato separazione delle sorgenti alla cieca (BSS). Uno degli scopi principali della BSS è quello di estrarre le sorgenti indipendenti da una mistura di segnali. Allo scopo di spiegare meglio il problema consideriamo una serie di segnali emessi da diverse sorgenti fisiche e assumiamo che ci siano dei sensori in differenti posizioni, questi sensori registreranno una mistura dei segnali sorgente con intensità differente. Assumiamo inoltre che il mescolamento sia istantaneo, ovvero vengono trascurati gli sfasamenti tra i segnali sorgente. Formalmente il modello di mistura può essere scritto come:

$$\mathbf{x} = \mathbf{A} \mathbf{s} + \mathbf{n}$$

dove \mathbf{x} è un vettore m -dimensionale di osservazioni, \mathbf{s} è un vettore n -dimensionale le cui componenti sono mutuamente indipendenti, \mathbf{A} è una matrice $m \times n$ i cui elementi sono da stimare e \mathbf{n} rappresenta un rumore additivo. Generalmente questo ultimo termine viene ommesso in quanto può essere incorporato nel vettore \mathbf{s} . Si assume inoltre che il numero delle differenti misture m sia almeno uguale al numero delle sorgenti indipendenti n . In genere m è noto e spesso $m = n$. Solo uno dei segnali sorgenti s_i può essere di tipo gaussiano, in quanto non è possibile separare due o più sorgenti gaussiane (Bell and Sejnowski, 1995; Karhunen, 1996). Usando tecniche iterative, una matrice di separazione \mathbf{B} viene aggiornata in modo che $\mathbf{y} = \mathbf{B} \mathbf{x}$ sia la migliore stima possibile dei segnali sorgente \mathbf{s} .

Nel seguito verrà usato l'algoritmo a punto fisso FastICA (Hyvarinen et al., 2001). La regola di apprendimento FastICA trova una direzione, cioè un vettore unitario \mathbf{w} , tale che la proiezione $\mathbf{w}^T \mathbf{x}$

massimizza l'indipendenza delle singole sorgenti stimate \mathbf{y} . L'indipendenza viene generalmente misurata attraverso la cosiddetta *negentropy* definita come

$$J(\mathbf{x}) = H(\mathbf{x}_{\text{gauss}}) - H(\mathbf{x})$$

dove H è l'entropia di Kolmogorov

$$H = -\int p(\mathbf{x}) \log[p(\mathbf{x})] d\mathbf{x}$$

ed il suffisso *gauss* viene riferito ad una variabile di tipo gaussiano.

Nell'algoritmo FastICA la *negentropy* viene approssimata dalla relazione

$$J_G(\mathbf{w}) = [E\{G(\mathbf{w}^T \mathbf{x})\} - E\{G(\nu)\}]^2$$

dove ν è una variabile gaussiana standard e $G(\bullet)$ è una funzione non lineare che qui viene posta come $G(y) = \log(\cosh(y))$. Va sottolineato che l'algoritmo richiede uno "sbiancamento" dei dati. Ovvero il vettore delle variabili osservabili \mathbf{x} deve subire la trasformazione $\mathbf{v} = \mathbf{Q}\mathbf{x}$ tale che $E\{\mathbf{v}\}=0$ e $E\{\mathbf{v}^T \mathbf{v}\}=1$. L'algoritmo FastICA unitario otterrà quindi (Hyvarinen et al., 2001)

$$\mathbf{w}^* = [E\{\mathbf{v}g(\mathbf{w})_i^T \mathbf{v}\} - E\{\mathbf{v}g'(\mathbf{w})_i^T \mathbf{v}\}] w_i$$

$$\mathbf{w}_i = \mathbf{w}_i^* / \|\mathbf{w}_i^*\|$$

dove $g(\bullet)$ è la derivata di $G(\bullet)$, nel nostro caso $g(y)=\tanh(y)$, e g' è la sua derivata rispetto ad y .

Tale algoritmo stima una sola delle componenti indipendenti. Per stimare le altre è necessario utilizzare molte unità FastICA (neuroni) con pesi $w_1 \dots \dots \dots w_n$. Allo scopo di evitare che differenti vettori convergano sullo stesso massimo è necessario decorrelare i risultati. L'algoritmo complessivo è di tipo iterativo e viene fermato quando la differenza tra i pesi di due iterazioni successive è inferiore ad un assegnato valore di soglia.

Analisi dei dati

In questa nota noi analizziamo i dati registrati alla rete GPS dell'INGV – Osservatorio Vesuviano di Napoli. Qui a causa della presenza di tre vulcani (Vesuvio, Campi Flegrei e isola di Ischia) e della elevata densità abitativa il rischio vulcanico è molto alto. Nel periodo che va dal 2001 all'inizio del 2003 la rete di stazioni GPS permanente è composta da 15 stazioni distribuite nelle tre differenti aree vulcaniche (fig. 1). La serie temporale si ferma all'inizio del 2003 quando la stazione di riferimento è cambiata a causa di problemi tecnici e logistici. La gestione della rete è stata programmata per delle soluzioni giornaliere in tempo quasi reale ed i dati sono stati elaborati con le procedure standard per reti di piccole dimensioni (Bottiglieri et al., 2007).

La posizione delle stazioni è stata calcolata usando POSI (vedi fig. 1) come stazione di riferimento. In questa nota abbiamo ignorato la stazione di Ischia che ha operato durante un periodo più breve il che riduce il numero di stazioni ad 11. In questo lavoro noi analizziamo il vettore spostamento composto il che equivale a prendere in considerazione la lunghezza della *baseline*. Questa misura ha il vantaggio di essere invariante per rotazione e traslazione del sistema di riferimento. Inoltre un test iniziale aveva mostrato che i risultati non cambiano in maniera significativa se si usano le differenti componenti dello spostamento del suolo (Bottiglieri et al., 2007).

Applicando l'ICA ai nostri segnali si ottengono, come componenti indipendenti, 5 segnali periodici con periodo pari a 341, 146, 93, 15 e 5 giorni rispettivamente (fig. 2 colonna di sinistra). L'incertezze sui pesi è rappresentata dalla tolleranza della procedura iterativa che qui è stata posta pari a 10^{-8} .

Ovviamente la natura periodica delle componenti indipendenti è stata stabilita tramite una usuale analisi spettrale (fig.2 colonna di destra). Alcune di queste componenti possono essere attribuite ad effetti mareali, mentre per altre non si hanno certezze sulla loro natura (Dong et al., 2002). Si noti che l'ICA è uno strumento di analisi molto più efficace dell'analisi spettrale tradizionale. Questa infatti non sempre riesce a riconoscere componenti periodiche individuate dall'ICA (De Lauro et al., 2005). Si veda ad esempio lo spettro del segnale registrato alla stazione di ACAE (fig. 3).

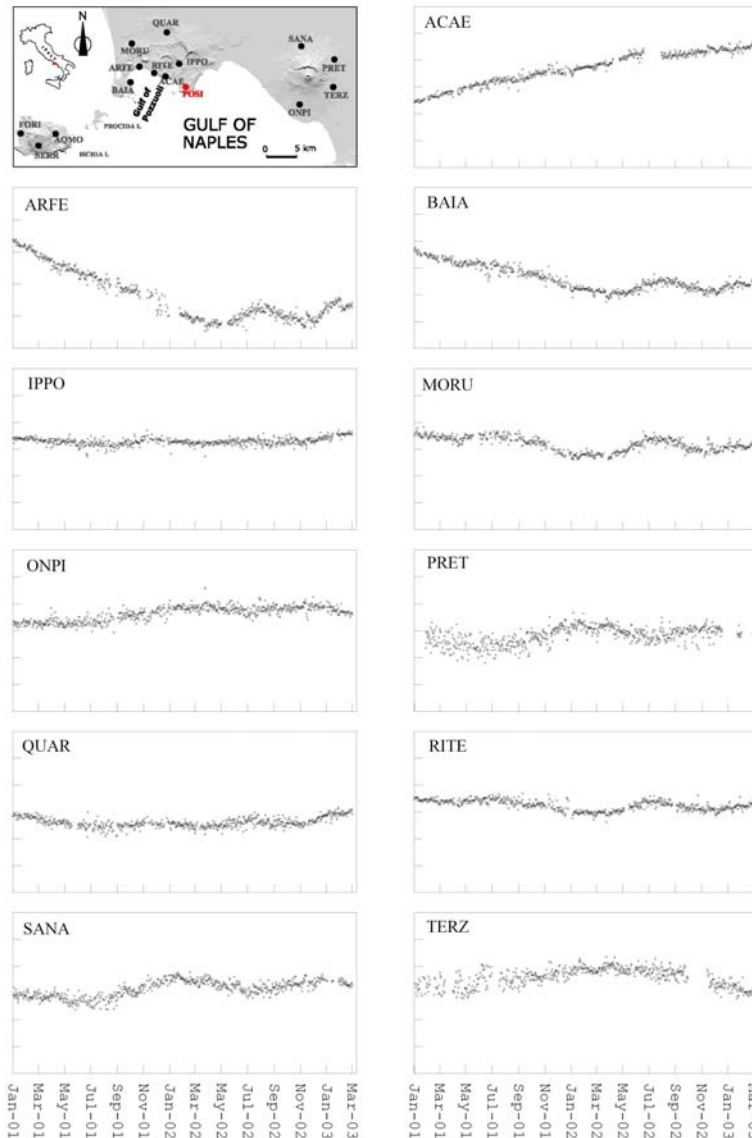


Figura 1 - Mappa delle stazioni permanenti GPS utilizzate e serie temporali delle baseline. Ogni divisione sull'asse verticale rappresenta 1 cm.

Come si è già osservato, l'ICA è in grado di separare i fattori indipendenti che soggiacciono ad un set di dati multivariati. Ciò implica che i segnali riconosciuti come componenti indipendenti rappresentano il livello di deformazione di background dell'intera area, essendo il "rumore" comune a tutte le stazioni. Il segnale residuo rappresenta invece il rumore locale stazionario a cui si aggiungono eventuali segnali transienti. Sfortunatamente l'ICA non è in grado di valutare l'ampiezza relativa dei segnali separati, cosicché sembrerebbe impossibile valutare il livello di background. D'altro canto le componenti individuate sono tutte periodiche e stazionarie. Pertanto la presenza di eventuali segnali transienti può essere evidenziata da una semplice analisi di stazionarietà.

Una serie temporale è detta stazionaria all' n -esimo ordine, se il suo momento statistico di ordine n non dipende dal tempo. Nel nostro caso una deformazione transiente può essere facilmente riconosciuta in quanto essa perturba significativamente il valor medio di un segnale stazionario. Un test molto utile per verificare l'eguaglianza tra due valori medi è rappresentato dal test di Student.

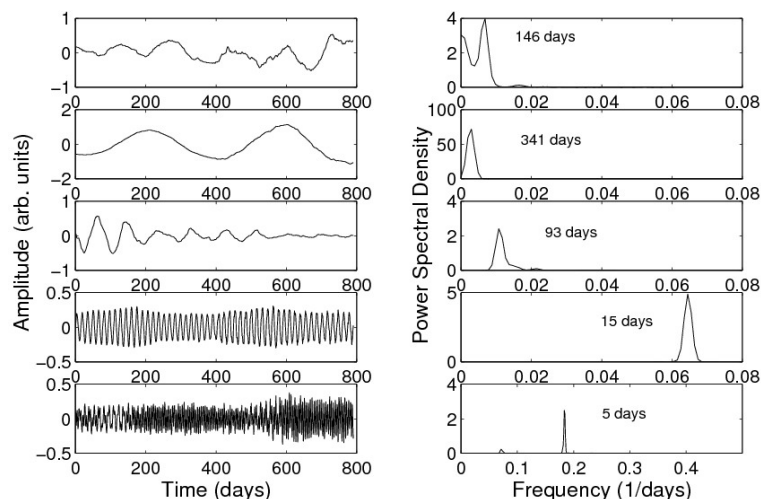


Figura 2 – Le componenti indipendenti estratti dai segnali con la tecnica ICA (sinistra) ed i corrispondenti spettri (a destra).

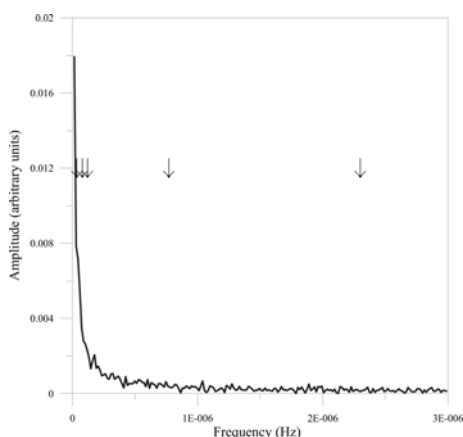


Figura 3 – Spettro della serie temporale registrata ad ACAE (riportata come esempio). Le frecce indicano le frequenze estratte dall'ICA (vedi fig. 2).

La stazionarietà è stata quindi verificata utilizzando un test t a due code con un livello di confidenza del 95%. I segnali sono stati separati in finestre non sovrapposte di 21 giorni. Tale scelta è dettata da un compromesso tra il numero di punti necessari ad avere una buona stima del valor medio ed una lunghezza sufficientemente breve da consentire un buon campionamento della serie. Il t test è stato applicato a tutte le coppie possibili di finestre. La serie è stata considerata stazionaria se il 98% delle coppie superava il test. Tale scelta è scaturita da un test su 500 realizzazioni di segnali stazionari per i quali si osservava un numero di coppie di finestre che non superavano il test pari al 2%. Come risultato si è ottenuto che le stazioni ACAE ed ARFE, posizionate rispettivamente ad ovest ed est della zona di massima inflazione, mostrano variazioni significative della linea base. La prima stazione rivela una estensione della linea base di circa 1 cm per anno, mentre per la seconda si osserva una contrazione di circa 1.6 cm per anno. Anche la stazione BAIA mostra, almeno fino a tutto il 2001, un'evidente contrazione, ma minore ed al limite del livello di stazionarietà prescelto.

Conclusioni

Abbiamo analizzato i dati di deformazione registrati alla rete GPS dell'area napoletana allo scopo di verificare la possibilità di utilizzare una nuova tecnica di analisi. Dapprima abbiamo eseguito una analisi delle componenti indipendenti in modo da separare i segnali sorgente che compongono le registrazioni alle stazioni della rete. Tali componenti rappresentano il livello di deformazione di background a scala regionale. I 5 segnali periodici possono essere considerati stazionari, mentre il

rumore locale di background può essere assunto come stazionario durante il periodo considerato. Si è quindi proceduto a valutare la stazionarietà dei segnali tramite una analisi della media mobile lungo i segnali registrati. L'applicazione di tale tecnica ha consentito di affermare che le stazioni alle quali si ha una variazione della distanza significativa, sono ACAE ed ARFE. La variazione riscontrata a BAIA, invece, è da considerare al limite della significatività, con i parametri scelti. Vogliamo infine sottolineare che la tecnica innovativa qui utilizzata può essere proficuamente usata per migliorare la sorveglianza delle aree vulcaniche.

Ringraziamenti

Questa ricerca è stata parzialmente finanziata con i fondi dell'UR V3_4/13 nell'ambito della Convenzione INGV-DPC 2004-2006.

Bibliografia

- Bell A. J. and Sejnowski T. J. (1995). An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129-1159.
- Bonforte A., Guglielmino F., Palano M., Puglisi G. (2004). A syn-eruptive ground deformation episode measured by GPS during the 2001 eruption on the upper Southern flank of Mt Etna. *Bull. Volcan.*, 66, 336-341.
- Bonforte A. and Puglisi G. (2006). Dynamics of the eastern flank of Mt. Etna volcano (Italy) investigated by a dense GPS network. *J. Volcan. Geotherm. Res.*, 153, 357-369.
- Czajewski J. (2004) The accuracy of the global positioning system. *IEEE Instru. Meas. Mag.*, 7, 56-60.
- Bottiglieri M., Falanga M., Tammara U., Obrizzo F., De Martino P., Godano C. and Pingue F. (2007). Independent component analysis as a tool for ground deformation analysis. *Geophys. J. Int.* 168, 1305-1310, doi:10.1111/j.1365-246X.2006.03264.x.
- De Lauro E., De Martino S., Falanga M., Ciaramella A., Tagliaferro R. (2005). Complexity of time series associated to dynamical systems inferred from independent component analysis. *Phys Rev E* 72,46712-46725.
- Dong D., Fang P., Bock Y., Cheng M.K. and Miyazaki S. (2002). Anatomy of apparent seasonal variations from GPS-derived site position time series. *J. Geophys. Res.*, 107, B4, 10.10129/2001JB000573.
- Fernandez J., Gonzales-Matesanz F. J., Prieto J. F., Rodriguez-Velasco G., Staller A., Alomso-Medina A. and Charco M. (2004). GPS monitoring in the N-W part of the volcanic island of Tenerife, Canaries, Spain: strategy and results. *Pure Appl. Geophys.*, 161, 1359-1377.
- Hyvarinen A., Karhunen J., Oja E. (2001). Independent Component Analysis. John Wiley & Sons.
- Karhunen J. (1996). Neural approach to independent component analysis and sources separation in *Proceedings of Fourth European Symposium on Artificial Neural Networks* , 249-266
- Lowry A.R., Hamburger M.W., Meertens C.M., Ramos E.G. (2001). GPS monitoring of crustal deformation at Taal Volcano, Philippines, *J. Volc. Geother. Res.*, 105, 35-47.
- Pingue F., Del Gaudio C., De Natale G., Obrizzo F., Sepe V., Cecere G., De Martino P., Malaspina S., Serio C., Siniscalchi V., Tammara U. (2003). Monitoring and deformation analysis in neapolitan area (Southern Italy) in *Proceedings of 11th Int. FIG Symposium on Deformation Measurements, Santorini-Greece*, 105-116.
- Savage J.C. (1988). Principal Component analysis of geodetically measured deformation in Long Valley caldera. eastern California, 1983-1987. *J. Geophys. Res.*, 93, B11, 13297-13305.
- Silver P.G., Bock Y., Agnew D., Henyey T., Linde A.T. , McEvelly T.V., Minster J.-B., Romanowicz B.A. , Sacks I. S., Smith R. B., Solomon S.C. , Stein S.A. (1999). A Plate Boundary Observatory, *IRIS Newsletter*, XVI, 2, 7-9.