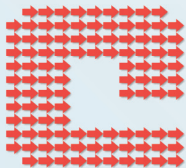


DOiT: UNO STRUMENTO "USER-FRIENDLY" PER L'ANALISI DI DATI GEODETICI



di Stefano Tagliaventi

Il Data Optimization integrated Tool (DOiT) è uno strumento completo che integra tecnologie web, sistemi e strutture *HPC* per l'analisi di problemi relativi alla geofisica dove l'elevata mole di dati e la complessità computazionale dei problemi in analisi ne rendono fondamentale l'utilizzo.

Dott. Stefano Tagliaventi
CASPUR
Gruppo di Matematica
Computazionale e Applicazioni
s.tagliaventi@caspur.it

• Abstract

The Data Optimization integrated Tool (DOiT) is part of a collaboration between CASPUR and the Italian national institute of geophysics and volcanology INGV. DOiT is a tool that mix together web technologies, systems and High Performance Computing (HPC) for the analysis of geophysical problems. One of the main feature of DOiT is the user interface that with the use of new web tools makes it very similar to a desktop application with the advantage of the high portability and the ease of access.

The main goal of this project is to make HPC systems usable in a very easy way in geophysical research, where the computational complexity and the high size of the problems make its use a must.

Il progetto DOiT è il risultato di una collaborazione tra il CASPUR e l'Istituto Nazionale di Geofisica e Vulcanologia, il cui scopo è quello di fornire uno strumento completo e di semplice utilizzo per l'analisi di problemi relativi alla geodesia. DOiT in particolare è rivolto ad una classe di problemi che possono essere affrontati utilizzando un metodo statistico denominato "metodo di inversione" dei dati, che utilizza come osservabili le deformazioni della superficie terrestre. Di questa classe fanno parte soprattutto problemi riguardanti la vulcanologia e ciò che riguarda le deformazioni del terreno dovute a processi sismici. Più in generale il campo di interesse sono quei problemi che hanno come obiettivo quello di determinare quantità di interesse (tipicamente in geofisica sono le proprietà della terra in profondità) che sono solo indirettamente collegate alle osservazioni superficiali. In questo scenario si inserisce la modellistica geofisica in quanto ad ogni problema inverso corrisponde (necessariamente) un problema "diretto" detto *forward-problem*, rappresentato da un modello fisico con il quale si effettuano previsioni sullo stato del sistema reale e con il quale vanno confrontati i dati sperimentali. L'idea fondamentale del metodo di inversione è di utilizzare le informazioni fornite dalla differenza tra i dati sperimentali e le previsioni del modello per intervenire iterativamente sulle incognite del modello stesso. Se la relazione tra i dati e le inco-

gnite del modello è non lineare il problema può essere molto difficile da risolvere.

Tale metodo è una procedura iterativa che può convergere ad una soluzione che rappresenti i dati attraverso il modello scelto. All'inizio del suo utilizzo in campo geofisico, negli anni '70, il metodo di inversione era associato ad approssimazioni lineari di modelli non-lineari utilizzati a causa della difficoltà di trattare in modo analitico tali non-linearità. Ora grazie alla crescente disponibilità di risorse di calcolo sempre più elevate è diventato possibile analizzare tali problemi nella loro totale non-linearità.

Da questo punto di vista DOiT rappresenta una soluzione a questo tipo di esigenza, rendendo di fatto molto semplice l'utilizzo delle risorse di calcolo ad alte prestazioni. Il cuore del metodo di inversione utilizzato in DOiT è l'algoritmo di inversione *NA (Neighbourhood Algorithm)* ideato e sviluppato da Malcolm Sambridge [1] e che consiste in quattro passaggi principali:

1. generazione di un insieme uniforme di n modelli nello spazio dei parametri;
2. calcolo della funzione di *misfit* (stima della differenza tra valori osservati e teorici) per l'insieme di valori generati e selezione di un sottoinsieme con i valori più bassi;
3. generazione di nuovi modelli effettuata attraverso una ricerca di tipo *random-walk* nelle celle di Voronoi (che equivale ad un metodo di ripartizione dello spazio) corrispondenti al sottoinsieme prima selezionato;
4. ritorno al passo 2 fino all'eventuale convergenza della funzione di *misfit*.

L'algoritmo termina quando è stata raggiunta la precisione voluta, quando cioè la differenza tra i valori determinati a partire dal modello in esame differiscono, al massimo di una quantità predefinita, dai dati osservati. L'algoritmo di inversione che è scritto in FORTRAN90 viene eseguito, su un sistema cluster ed è stato parallelizzato utilizzando MPI (una libreria che permette la programmazione parallela). L'esecuzione del processo richiede la generazione dei files di configurazione e la sottomissione del job in un sistema di code. Le configurazioni principali che è possibile scegliere sono:

- **Il tipo di modello diretto da utilizzare** (la scelta del modello dipende dal problema geofisico in esame, può essere ad esempio un problema di tipo vulcanologico, di tipo sismico o entrambi). In DOiT sono stati implementati tre diversi modelli fisici, un modello ad elementi finiti, un modello che rappresenta la combinazione lineare dei tre modelli base, e un modello dove la deformazione della crosta è dovuta ad eventi sismici. Ogni modello è caratterizzato da un numero di parametri che nel caso della combinazione lineare è variabile.

- **L'algoritmo di inversione da utilizzare.** In DOiT, sono stati implementati (oltre all'algoritmo NA) metodi di ricerca di tipo Montecarlo e NA-Bayes. La selezione non si limita pertanto alla scelta dell'algoritmo, ma implica la configurazione di parametri con i quali l'algoritmo deve essere eseguito.

- **L'insieme dei dati su cui effettuare l'inversione.** I dati che rappresentano gli spostamenti della crosta terrestre sono quelli più utilizzati in campo geofisico



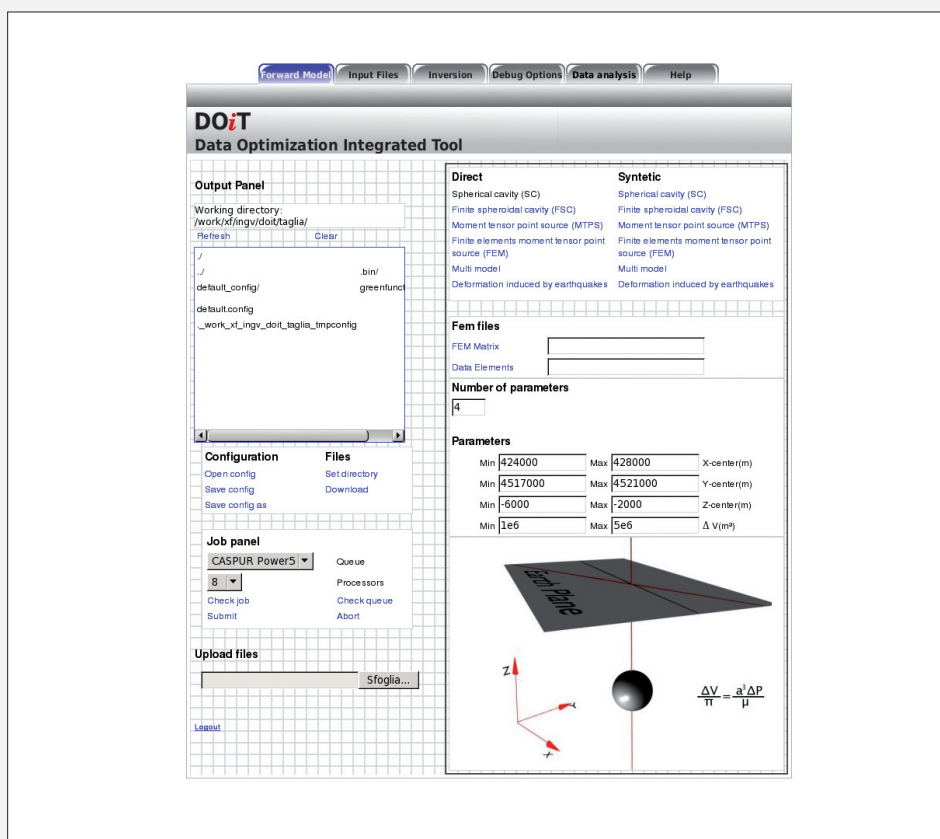
e sono di tipo GPS (Global Positioning System), EDM (Electronic Distance Meter) e SAR (Synthetic Aperture Radar) con le varianti InSAR e DInSAR¹. È stata inoltre inserita la possibilità di effettuare l'inversione su un *dataset* che rappresenta una combinazione lineare dei dati sopracitati.

- **L'ambiente di test e verifica.** È possibile far generare a DOiT insiemi di dati che verranno utilizzati per verificare la convergenza e precisione dell'algoritmo di inversione rispetto al modello scelto.

La procedura di configurazione del problema e di sottomissione al sistema di *code* del *cluster* viene effettuata in modo molto intuitivo con l'interfaccia web di DOiT. Ogni utente ha a disposizione una directory dove eseguire i propri lavori. Tale directory è parte di un *file-system* distribuito e viene gestita dall'utente attraverso l'interfaccia, grazie ad un *file-manager* realizzato *ad hoc*.

Figura 1

Interfaccia web di DOiT.



DOiT si caratterizza pertanto per una struttura piuttosto articolata che fa uso di diverse tecnologie web che permettono l'integrazione con i sistemi sottostanti: un server di gestione dell'intera catena di eventi (che agisce come un *application-server*) scritto in C++ ed un database per la gestione delle sessioni e delle configurazioni dell'algoritmo di inversione.

Ha inoltre una struttura modulare che permette in modo agevole l'implementazione di nuovi blocchi e di nuove procedure all'interno dei singoli blocchi (in fi-

¹ InSAR: Interferometric Synthetic Aperture Radar; DInSAR: Differential Interferometric Synthetic Aperture Radar.

gura è mostrata l'interfaccia *web-like* dell'ambiente DOiT). Tutte le comunicazioni tra i blocchi distinti (anche all'interno della stessa macchina) avvengono con l'utilizzo di comunicazioni TCP/IP incapsulate nel protocollo SSL. È possibile in questo modo localizzare i diversi blocchi su macchine differenti mantenendone sicura la comunicazione. Relativamente alle sue applicazioni "sul campo", DOiT è stato utilizzato con successo per l'analisi dei dati del monte Etna (1993-1997) e per l'analisi del sollevamento dei Campi Flegrei del 2004-2006 [2].

• Bibliografia

-
- [1] Sambridge, M. (1999). *Geophys. J. Int.*, **138**, 479-494; 727-746.
- [2] Trasatti, E., *et al.* (2008). The 2004–2006 uplift episode at Campi Flegrei caldera (Italy): Constraints from SBAS-DInSAR ENVISAT data and Bayesian source inference. *Geophys. Res. Lett.*, **35**, L07308, doi:10.1029/2007GL033091.