

Rapporti tecnici

INGV

**Porting su Grid del modello TEPHRA
per la previsione della dispersione
delle ceneri vulcaniche**

105



Direttore

Enzo Boschi

Editorial Board

Raffaele Azzaro (CT)

Sara Barsotti (PI)

Mario Castellano (NA)

Viviana Castelli (BO)

Anna Grazia Chiodetti (AC)

Rosa Anna Corsaro (CT)

Luigi Cucci (RM1)

Mauro Di Vito (NA)

Marcello Liotta (PA)

Lucia Margheriti (CNT)

Simona Masina (BO)

Nicola Pagliuca (RM1)

Salvatore Stramondo (CNT)

Andrea Tertulliani - coordinatore (RM1)

Aldo Winkler (RM2)

Gaetano Zonno (MI)

Segreteria di Redazione

Francesca Di Stefano - coordinatore

Tel. +39 06 51860068

Fax +39 06 36915617

Rossella Celi

Tel. +39 06 51860055

Fax +39 06 36915617

redazionecen@ingv.it



Rapporti tecnici INGV

PORTING SU GRID DEL MODELLO TEPHRA PER LA PREVISIONE DELLA DISPERSIONE DELLE CENERI VULCANICHE

Gaetano Spata¹, Antonio Drago², Fabrizio Pistagna², Simona Scollo¹, Mauro Coltelli¹,
Michele Prestifilippo¹, Danilo Reitano¹

¹INGV (Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania)

²COMETA (Consorzio Multi Ente per la promozione e l'adozione di Tecnologie di calcolo Avanzato)

105

Indice

Introduzione	5
1 Le ceneri vulcaniche ed i rischi ad esso associati	7
2 Il modello TEPHRA per la previsione della dispersione delle ceneri vulcaniche	8
3 L'utilizzo del modello TEPHRA presso l'INGV, sezione di Catania	9
4 Il progetto PI2S2 e il paradigma Grid	11
5 Porting del modello TEPHRA su Grid	12
6 Risultati e conclusioni	14
Bibliografia	17

Introduzione

Durante gli ultimi anni, l'Etna ha mostrato un aumento dell'attività esplosiva [Branca and Del Carlo, 2005] con la formazione di colonne eruttive alte fino ad una decina di chilometri al di sopra del livello del mare. La cenere vulcanica prodotta ha creato notevoli disagi alla popolazione, come per esempio la chiusura dell'Aeroporto Internazionale di Catania, posto solo a circa 30 km dai crateri sommitali. Notevoli danni, per esempio, furono causati dall'attività prolungata del 2002-03 [Andronico et al., 2008] quando si ebbero ben 58 giorni di continua emissione di cenere costringendo alcuni comuni del territorio etneo a chiedere lo stato di calamità naturale. Per ridurre i danni provocati dalla cenere vulcanica è necessario fornire alla Protezione Civile informazioni precise sull'area interessata dalla presenza di cenere vulcanica come anche sulla quantità di cenere presente. L'Istituto Nazionale di Geofisica e Vulcanologia (INGV) ha realizzato negli ultimi anni un sistema multidisciplinare di previsione e monitoraggio delle ceneri vulcaniche atto a fornire in tempi rapidi alle autorità competenti tali informazioni. Questo sistema si basa su l'utilizzo di diversi modelli di previsione e su una serie di strumenti capaci di misurare la caduta e la dispersione di cenere vulcanica [Coltelli, 2006].

In letteratura sono presenti diversi modelli di dispersione in grado di simulare la caduta e la dispersione della cenere vulcanica in atmosfera. Tali modelli risolvono l'equazione di avvezione e diffusione utilizzando approcci di tipo euleriano o lagrangiano. Modelli euleriani sono per esempio ASHFALL [Macedonio et al., 1988], FALL3D [Costa et al., 2006] sviluppati dall'Osservatorio Vesuviano, MEDIA [Sandu et al., 2003] and CANERM [D'Amours, 1998] utilizzati dal Volcanic Ash Advisor Center (VAAC) di Tolosa and Monreale, ovvero centri che hanno la responsabilità di coordinare e disseminare tutte le informazioni sulla presenza di cenere vulcanica per evitare che gli aerei attraversino la nube eruttiva prodotta durante l'attività esplosiva di un qualunque vulcano. Alcuni modelli lagrangiani sono invece il PUFF utilizzato all'Alaska Volcano Observatory [Searcy et al., 1998], il VOL-CALPUFF sviluppato dall'INGV di Pisa [Barsotti et al., 2008], HYSPLIT [Draxler e Hess, 1998], NAME [Ryall e Maryon, 1998] utilizzati dal VAAC di Londra (UK) e Washington (USA). Sotto semplificate assunzioni alcuni modelli come HAZMAP [Macedonio et al., 2005] e TEPHRA [Bonadonna et al., 2005] risolvono l'equazione di avvezione e diffusione semi analiticamente. Tali modelli sono molto utilizzati per produrre mappe di hazard [Cioni et al., 2003] perché i modelli girano in pochi secondi e permettono di eseguire un notevole numero di simulazioni in tempi brevi, facilitando l'utilizzo di approcci probabilistici.

In particolare TEPHRA [Bonadonna et al., 2005] è un modello attualmente utilizzato presso l'Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania (INGV-CT), per la previsione del deposito al suolo delle ceneri vulcaniche, in presenza di attività esplosiva all'Etna. Il modello TEPHRA, implementato in linguaggio ANSI-C con l'ausilio delle librerie MPI (*Message Passing Interface*), sfrutta i vantaggi computazionali offerti dal calcolo parallelo. Per questo motivo il modello è stato già utilizzato per la costruzione di mappe di hazard per il Tarawera, in New Zealand [Bonadonna et al. 2005], per l'Etna [Scollo et al., 2006] e per studi di analisi di sensibilità dei modelli di dispersione delle ceneri vulcaniche [Scollo et al., 2008]. TEPHRA è attualmente eseguito su un cluster Beowulf presente presso l'INGV-CT. Giornalmente vengono simulati tre scenari eruttivi i cui parametri fisici sono stati caratterizzati in dettaglio dall'analisi dei depositi piroclastici: l'eruzione del 22 luglio 1998 [Andronico et al., 1999] che ha prodotto una colonna eruttiva alta fino a 12 km, l'attività esplosiva prodotta tra il 21 e 24 luglio del 2001 [Scollo et al., 2007] in cui una colonna con un'altezza massima di 5-6 km è perdurata per circa 4 giorni (Figura 1) e l'eruzione 2002-03 in cui una prolungata attività esplosiva ha prodotto colonne eruttive con altezze variabili tra 3.3 e 7 km [Andronico et al. 2008]. In questo rapporto si è lavorato per poter eseguire e valutare le performance del modello su una diversa piattaforma di calcolo ad alte prestazioni, basata sul paradigma emergente della Grid (griglia computazionale). L'infrastruttura di calcolo verso cui è stata effettuata tale attività di *porting* è la Grid realizzata nell'ambito del progetto PI2S2 "Progetto per l'Implementazione e lo Sviluppo di una e-Infrastruttura in Sicilia basata sul paradigma della Grid (<http://www.pi2s2.it>) del Consorzio COMETA (Consorzio Multi Ente per la promozione e l'adozione di Tecnologie di calcolo Avanzato) (<http://www.consorzio-cometa.it>).

Il lavoro ha previsto un'attività di modifica e adattamento dell'applicazione, nonché una sua configurazione secondo le regole di sottomissione previste per l'infrastruttura di calcolo suddetta. Sono stati analizzati i risultati ottenuti e sono state confrontate le performance raggiunte con quelle già ottenute sul cluster locale presente presso l'INGV-CT. Sono state individuate ed affrontate, con l'ausilio dei *site-manager* della Grid, le problematiche, sia di quelle di tipo architetturale sia quelle specifiche

dell'applicazione, legate in particolare alle esigenze di brevità e predicibilità dei tempi di esecuzione del modello TEPHRA.

È da evidenziare che la stessa metodologia descritta in questo rapporto potrebbe essere effettuata con qualunque altro modello di dispersione che gira in parallelo, come ad esempio la piattaforma APOLLO [Folch et al., 2008] anch'essa utilizzata all'INGV-CT, ed ancora più in generale con modelli indirizzati a risolvere altre problematiche (e.g. flussi lavici).



Figura 1. Colonna di cenere alta circa 4.5 km vista da NNW . Foto scattata il 24 Luglio del 2001 durante le attività di monitoraggio dell'INGV-CT.

1. Le ceneri vulcaniche ed i rischi associati

L'insieme del materiale non consolidato (prodotti piroclastici) rilasciato da un vulcano nell'atmosfera e depositato al suolo durante la sua attività esplosiva è detto “*tephra*” e comprende frammenti di dimensioni e peso variabile che vanno dalle cosiddette bombe e blocchi (>64 mm) al materiale più fino come la cenere vulcanica (<2 mm).

In base alla loro natura i prodotti piroclastici si dividono in frammenti juvenili derivanti dal magma frammentato, in frammenti litici, derivati da rocce all'interno del condotto formatosi in seguito ad eruzioni precedenti, e in vari tipi di cristalli..

A causa della piccola dimensione della cenere vulcanica essa può essere trasportata nell'atmosfera per giorni od addirittura mesi [Rose et al., 2001] e può ricoprire un'ampia area di dispersione che assume spesso una forma ellittica lungo la direzione dei venti prevalenti.

La cenere vulcanica è causa di svariati rischi sia per la salute che per la sicurezza dei territori coinvolti, talvolta con ripercussioni sociali ed economiche non indifferenti. L'inalazione della cenere può causare problemi al sistema respiratorio, specialmente per i malati di asma o enfisema, o provocare irritazioni e abrasioni ai polmoni e agli occhi ed in alcuni casi tumori [Horwell e Baxter, 2006]. Per quanto riguarda invece la sicurezza i principali rischi riguardano il traffico aereo [Casedevall, 1994]. Infatti, le particelle di cenere vulcanica una volta aspirate dalle turbine dei velivoli, fondono nella camera di combustione, all'interno della quale si raggiungono temperature intorno a 1400° C. La successiva solidificazione sulle palette e sulle parti in movimento delle turbine compromette il loro funzionamento fino a provocarne il blocco totale. La cenere inoltre, essendo dura e molto abrasiva, può erodere la struttura dell'aeromobile e delle superfici di volo. In particolare l'abrasione dei vetri della cabina di pilotaggio può ridurre fino a livelli di estrema pericolosità la visibilità a lungo raggio del pilota. I radar meteorologici presenti sugli aeromobili, anche i più moderni, non riescono a rilevare le ceneri vulcaniche, pertanto la previsione e la tempestiva segnalazione della loro presenza in atmosfera, specialmente di notte e in condizioni di scarsa visibilità, rappresentano ancora gli unici veri mezzi di prevenzione.

Nel territorio etneo, studiato e monitorato dall'INGV-CT, tale genere di rischi è molto presente poiché a pochi chilometri dal vulcano (all'incirca 30 km) sorge l'aeroporto internazionale di Fontanarossa, una delle aerostazioni più trafficate d'Italia, spesso soggetto a caduta di cenere vulcanica dell'Etna (Figura 2).

Anche il traffico stradale è messo a dura prova dalle emissioni di cenere. Queste, infatti, possono danneggiare i motori degli autoveicoli e, producendo una riduzione della visibilità, possono creare grossi problemi alla viabilità ed essere causa di incidenti stradali [Barnard, 2004].



Figura 2. Cenere vulcanica sulla pista dell'aeroporto di Fontanarossa a Catania durante l'eruzione del 2002. Foto di Mauro Coltelli.

Oltre ai trasporti, le ripercussioni della dispersione di cenere riguardano molti altri settori. La deposizione di cenere causa una contaminazione di pascoli e colture, con un'inevitabile ricaduta sugli allevamenti e sull'agricoltura, nonché una contaminazione delle acque di superficie. A causa del fluoro e dei metalli pesanti contenuti nelle particelle, i bacini e i corsi d'acqua presenti nei territori coinvolti possono risultare inquinati o comunque alterati. L'intorbidimento, l'alterazione del PH e della temperatura, sono alcuni degli effetti che possono talvolta rivelarsi fatali per la fauna ittica.

L'accumulo di ceneri può causare il crollo dei tetti e danni alle coperture edili in genere. Il danno può essere amplificato dall'acqua piovana che combinata con la cenere causa l'intasamento delle grondaie con la conseguente infiltrazione nella struttura delle costruzioni.

In ultimo vanno tenuti presenti i disturbi che le particelle disperse nell'atmosfera provocano nella ricezione dei segnali elettromagnetici, e quindi nelle telecomunicazioni.

2. Il modello TEPHRA

TEPHRA [Bonadonna et al. 2005] è un modello di avvezione e diffusione che descrive la diffusione, il trasporto e la sedimentazione di particelle vulcaniche risolvendo analiticamente l'equazione della conservazione della massa. Le particelle vulcaniche di una specifica dimensione sono rilasciate da un punto sorgente ad una determinata altezza della colonna eruttiva. Attraverso l'equazione di avvezione e diffusione è possibile stimare in un punto specifico (x, y) la frazione di massa data da una particolare classe granulometrica delle particelle. Queste sono rilasciate istantaneamente dalla sorgente e sono assunte essere sferiche con una velocità di caduta che è funzione del numero di Reynolds e della stratificazione atmosferica. Le particelle sono soggette anche alla turbolenza dell'atmosfera e dell'influenza della nube gravitativa vulcanica e sono trasportate grazie all'azione dei venti. Il modello può anche utilizzare approcci probabilistici e quindi produrre mappe di probabilità e curve di *hazards*. I risultati ottenuti dal modello sono visualizzati sotto forma di mappe di dispersione che localizzano geograficamente il deposito delle ceneri vulcaniche o mappe di probabilità nel caso vengano simulati più eventi eruttivi.

Diversi fattori contribuiscono a rendere piuttosto complessi i calcoli effettuati dal modello TEPHRA. L'algoritmo richiede la soluzione dell'equazione di avvezione e diffusione in un punto (x, y) per ogni classe granulometrica considerata, che può essere a sua volta emessa da diversi punti della colonna eruttiva. Dal punto di vista computazionale, questi steps risultano piuttosto lenti. In una tipica simulazione, l'accumulo delle ceneri viene calcolata in punti posti a diversa distanza dal cratere. Lo spazio considerato dalle simulazioni viene suddiviso nei punti di una griglia: un'approssimazione di 5000 punti produce mappe di dispersione soddisfacenti per il dominio considerato nel caso dell'Etna.

L'oneroso carico computazionale del modello TEPHRA è in qualche modo ridotto grazie all'utilizzo del calcolo parallelo. Il parallelismo utilizzato dal codice di TEPHRA è di tipo *SPMD* (*Single Program - Multiple Data*). Istanze multiple dello stesso codice vengono eseguite parallelamente su diversi nodi (processori) con un set diverso di dati di input. Ciò è realizzabile perché i risultati ottenuti dall'applicazione del modello su un singolo punto della griglia non dipendono dagli altri. Pertanto l'intero set dei dati in input viene suddiviso fra i vari nodi di calcolo. Ogni nodo esegue la computazione vera e propria del modello con il proprio set parziale di input. Alla fine un nodo *master* si occupa di raccogliere ed ordinare l'insieme di risultati e di restituirli in output nella loro completezza, per la creazione delle mappe.

Gli input del modello TEPHRA sono: un file di configurazione del modello (*new_tephra.conf*), i parametri che identificano l'area geografica d'interesse (*etnapoints.in*), i parametri che rappresentano l'eruzione simulata (*eruption_strong.in*) e i dati meteorologici che danno le informazioni sulla direzione e velocità dei venti (*wind.dat*). Mentre i primi sono dati che restano fissi nell'arco della simulazione della stessa eruzione poiché funzione dello scenario eruttivo simulato, i dati riguardanti le previsioni dei venti variano in funzione del particolare giorno ed ora considerati.

3. L'utilizzo del modello TEPHRA presso l'INGV-CT

Il cluster dell'INGV-CT (Figura 3), su cui viene quotidianamente eseguito il modello TEPHRA, è costituito da 24 server Proliant DL 145 OPTERON 248. Ciascuna di queste macchine possiede due processori da 2.6 GHz/1Mb, 1 Gb di memoria RAM ed un HDD da 80 Gb. Solamente sul nodo master troviamo un ulteriore HDD da 160 Gb. I server sono collegati tra di loro attraverso un HP Procurve switch 2810-48G con 64 Mb di SDRAM e 16 Mb di memoria flash. Lo *switch* possiede 48 porte ad 1 Gb, permettendo di utilizzare entrambe le porte Ethernet di ogni server.

Sul nodo *master* è installato il sistema operativo SUSE Linux Enterprise Server, oltre ai vari compilatori ed alle librerie per i linguaggi C e Fortran, tra cui le librerie MPI per il calcolo parallelo. Tutti i server *slaves* invece lavorano creandosi un'immagine del server master, facendo il *boot* dalla rete tramite NFS.



Figura 3. Il Beowulf cluster presente all'INGV-CT.

Ogni giorno, con l'ausilio di un sistema automatizzato, a partire dai dati meteorologici forniti sia dall'Aeronautica Militare che dall'ARPA, il modello TEPHRA simula diversi scenari eruttivi, relativi rispettivamente alle eruzioni dell'Etna del 1998, del 2001 e 2002-03, i cui depositi sono stati analizzati in dettaglio [Andronico et al., 1999; 2008; Scollo et al, 2007].

I file contenenti i dati meteorologici che ci vengono forniti, sono in formato Grib, un particolare formato costituito da un numero di record, divisi a loro volta in sottosezioni.

Tramite degli script implementati all'interno dell'INGV-CT, dai file Grib vengono estrapolate le informazioni riguardanti l'intensità e la direzione dei venti, la temperatura, la pressione e l'umidità.

Ogni giorno si elaborano ben 16 simulazioni. Ciascuna di esse fornisce una previsione per un intervallo temporale di 3 ore, quindi, vengono prodotte le previsioni per le successive 48 ore e viene effettuata in meno di 100 s utilizzando un numero maggior di 40 processori (Figura 4).

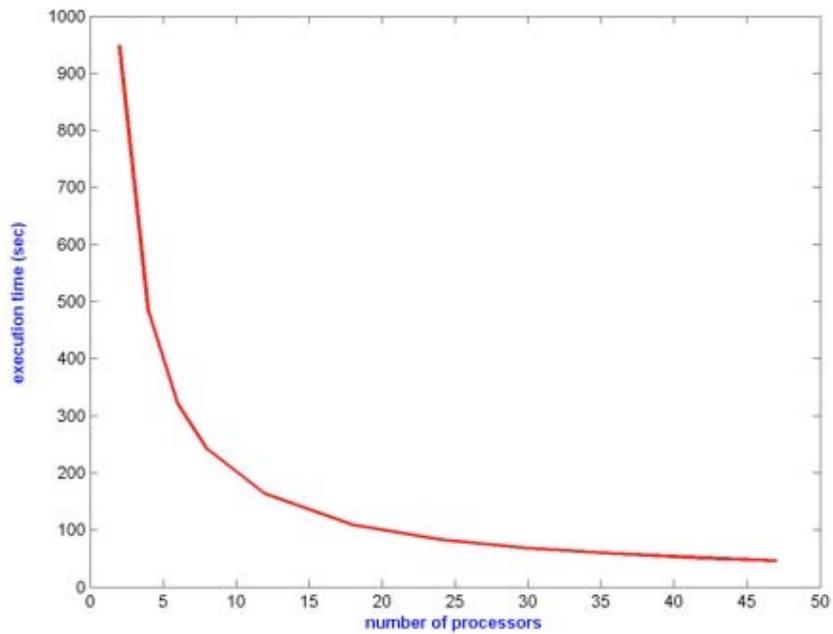


Figura 4. Tempi di elaborazione ottenuti sul cluster locale per una singola simulazione del modello TEPHRA al variare del numero di processori utilizzati.

Dai dati restituiti in output si ricavano delle mappe di ricaduta al suolo delle ceneri vulcaniche (Figura 5). Queste sono fornite entro la mattinata al Dipartimento della Protezione Civile che si preoccuperà di avvertire le diverse autorità aeroportuali, di lanciare eventuali allarmi e di prendere decisioni in merito a possibili misure cautelative da intraprendere ai fini della pubblica sicurezza.

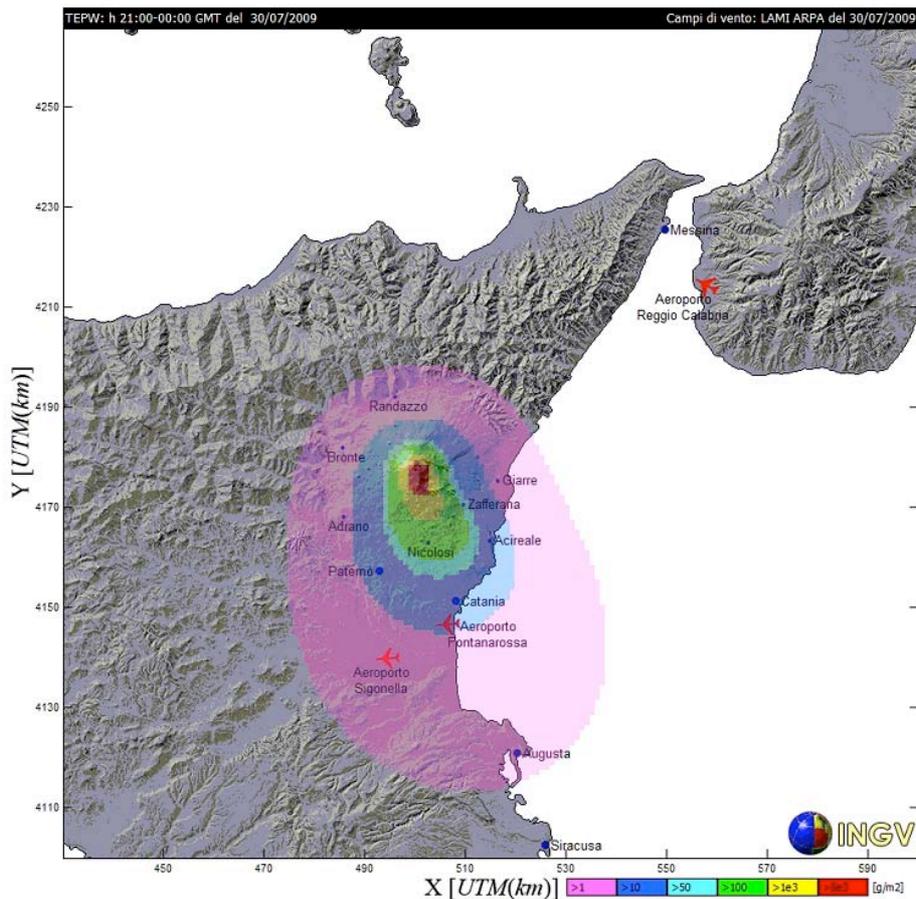


Figura 5. Esempio di una mappa della dispersione della cenere dell'Etna elaborata dal modello TEPHRA.

4. Il progetto PI2S2 e il paradigma Grid

L'INGV-CT, membro del consorzio COMETA (COnsorzio Multi Ente per la promozione e l'adozione di Tecnologie di calcolo Avanzato), insieme all'INFN, al Consorzio SCIRE e alle tre università statali siciliane, partecipa al progetto denominato PI2S2, "Progetto per l'Implementazione e lo Sviluppo di una e-Infrastruttura in Sicilia basata sul paradigma della Grid (Figura 6). Tale progetto, che rientra nel Programma Operativo Nazionale "Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione" 2000-2006 (PON Ricerca), si è proposto di dotare la Sicilia di una grande infrastruttura distribuita di calcolo e di storage al servizio della comunità scientifica e delle aziende. Questa struttura, di elevato livello tecnologico, si propone come un vero e proprio Laboratorio Virtuale capace di fornire i propri servizi "on demand" alle Università ed agli Enti di Ricerca coinvolti, nonché alle PMI (Piccole e Medie Imprese) che ne fanno richiesta. Ciò mira a favorire uno scambio di tecnologie e conoscenze e quindi un incremento della collaborazione scientifica tra Enti e Atenei, oltre ad una crescita di competitività per le aziende che usufruiscono della potenzialità della Grid.



Figura 6. Logo del progetto PI2S2.

Il paradigma di calcolo distribuito utilizzato è quello del *Grid Computing*: una "griglia" decentralizzata e delocalizzata di calcolatori interconnessi attraverso una rete (Internet) in grado di condividere le proprie risorse (CPU e memoria di storage) e fornirle potenzialmente a qualunque utente ad essa connesso. I "nodi" della Grid implementata per il progetto PI2S2 sono dislocati geograficamente tra i tre poli di Catania (Laboratori INAF, INFN e UNICT), Palermo (Polo per il Supercalcolo Scientifico) e Messina (Facoltà di Ingegneria dell'Università di Messina). La Grid siciliana è poi collegata anche al circuito della Grid europea, con cui condivide peraltro il *middleware*, GLite su piattaforma Unix-Linux, ovvero lo strato di software che gestisce tutti i servizi Grid di base.

Dopo una prima fase di progettazione, realizzazione e messa in opera dell'infrastruttura di calcolo attraverso l'installazione dell'hardware e del relativo *middleware*, il progetto è ad oggi nella sua fase finale, che prevede la realizzazione di software per l'interfacciamento con il *middleware* e il *porting* di applicazioni su Grid.

L'INGV-CT ha lavorato affinché alcune delle applicazioni scientifiche utilizzate per le proprie attività di ricerca potessero sfruttare la potenza della griglia computazionale di PI2S2, come descritto nel seguente paragrafo.

5. Porting del modello TEPHRA su Grid

La natura parallela dell'algoritmo utilizzato dal modello TEPHRA ha portato a testare l'esecuzione dell'applicazione su una potente infrastruttura di calcolo distribuita quale la Grid di PI2S2.

Il modello TEPHRA è scritto in linguaggio ANSI C, utilizza le librerie GC (ver.6.7 http://www.hpl.hp.com/personal/Hans_Boehm/gc/) distribuite con licenza *open source* dalla HP come ausilio per la gestione della memoria allocata (*malloc*) e sfrutta le specifiche di interfaccia per il calcolo parallelo MPI (*Message Passing Interface*) ver.1, standard de facto per la realizzazione di codice parallelo in C o Fortran, basato sullo scambio di messaggi tra processori secondo il paradigma *Send - Receive*.

Compilatori GCC, INTEL e PGI per programmi in linguaggio C sono presenti su tutti i nodi della Grid, e in particolare sui nodi di User Interface, quelli cioè a cui gli utenti hanno direttamente accesso, e che sono a loro volta i punti unici da cui poter accedere a tutti i servizi Grid, secondo un rigido controllo d'accesso di tipo PKI (*Public Key Infrastructure*), basato su un sistema di certificati digitali e crittografia a chiave pubblica.

Riguardo a MPI, la griglia di PI2S2 supporta due implementazioni *open-source*: MPICH, che implementa lo standard MPI 1.1, ed MPICH2, che implementa il più recente MPI 2.0.

Al momento il parallelismo realizzato attraverso MPI su Grid non è di tipo "geografico", ma opera solo nell'ambito dello stesso sito di calcolo (*Computing Element*), ciò significa che lo scambio di messaggi è permesso solo tra i processori dei nodi (*Worker Node*) dello stesso sito. Le reti disponibili per lo scambio di dati tra i processori sono di tipo GigaBit Ethernet o InfiniBand, la nuova tecnologia di rete a bassa latenza.

Durante la fase di *porting* su architettura Grid si sono utilizzati diversi approcci per ottenere la migliore stabilità ed efficienza in fase di *runtime*.

Per poter essere eseguito sul *middleware* della Grid, il modello TEPHRA è stato compilato attraverso un'interfaccia "User Interface", la cui configurazione di sistema è del tutto identica a quella dei *Worker Node* su cui l'applicativo è poi eseguito. Sono stati utilizzati compilatori Gnu GCC e Portland PGI.

Secondo le regole di funzionamento del paradigma Grid, ogni applicazione può essere eseguita sull'infrastruttura di calcolo previa sottomissione di un "job", opportunamente descritto in uno script file (Figura 7) secondo le direttive del linguaggio JDL (*Job Description Language*). Mediante questo file JDL è possibile definire tutta una serie di parametri e configurazioni necessarie per l'esecuzione dell'applicazione. In particolare vengono specificati i file di input e di output prodotti, eventuali requisiti richiesti per la sottomissione e, nel caso di job di tipo MPI, l'implementazione utilizzata e il numero di nodi richiesti.

```
[adrago@infn-ui-01 tephraJob]$ cat tephra.jdl
Type = "Job";
JobType = "MPICH";
Executable = "new_tephra";
NodeNumber = 30;
Arguments = "new_tephra.conf eruption_strong.in etnapoints.in wind00.in";
StdOutput = "stdout.out";
StdError = "stderr.out";
InputSandbox = {"mpi.pre.sh", "mpi.post.sh", "new_tephra", "wind00.in",
                "eruption_strong.in", "etnapoints.in", "new_tephra.conf"};
OutputSandbox = {"stdout.out", "stderr.out", "tephraoutput.out"};
RetryCount = 3;
Requirements = (
    other.GlueCEUniqueID=="infn-ce-01.ct.pi2s2.it:2119/jobmanager-lcglsf-long"
    || other.GlueCEUniqueID=="infn-ce-01.ct.pi2s2.it:2119/jobmanager-lcglsf-short"
);
```

Figura 7. Esempio di JDL utilizzato per la sottomissione su Grid del modello TEPHRA.

Va lanciato un job differente per ciascuna simulazione, passando in input ed a ciascun job, un set diverso di dati meteorologici in base alla fascia oraria che si vuole simulare.

Sono state adottate diverse strategie di sottomissione dei job, per cercare di ottimizzare i tempi di esecuzione del modello in base al carico di lavoro. Si è scelto di sfruttare la disponibilità di risorse di calcolo della Grid programmando, tramite script di *shell bash* opportunamente creati, il lancio simultaneo su job differenti di tutte le 16 simulazioni giornaliere.

Dopo una prima fase di sperimentazione, se da un lato i singoli tempi di esecuzione si sono rilevati competitivi rispetto a quelli già ottenuti su cluster locale della sezione INGV-CT, dall'altro, per via dei tempi di attesa delle code di job, i tempi globali sono risultati non soddisfacenti, in quanto talvolta eccessivamente lunghi e comunque non predicibili.

I sistemi presenti nella Sala Operativa dell'INGV-CT, devono operare *full time* H24 per garantire sempre le necessarie informazioni quando esse siano richieste. In quest'ottica durante periodi di quiete, ma soprattutto in regime di emergenza, risulta fondamentale per scopi sociali e non solo di ricerca, poter disporre dei risultati delle simulazioni nel più breve tempo possibile. Tematiche di *redundancy*, di QoS e tempestività dello svolgimento dei processi sono esigenze irrinunciabili se si vuole utilizzare la struttura del Grid a fini di monitoraggio e sorveglianza.

Per far fronte al problema della criticità dei tempi in cui deve essere eseguito TEPHRA, dovuta all'esigenza, spiegata sopra, di comunicare gli output elaborati alla Protezione Civile in tempi rapidi, è stato realizzato e sperimentato l'uso di code dedicate di job ad alta priorità (*Emergency Queue*). Questo servizio, riservato solo a determinati utenti con opportuni permessi, permette di bloccare temporaneamente tutti i job in esecuzione su un determinato sito della Grid affinché un job ad alta priorità, nella fattispecie il modello TEPHRA, possa andare immediatamente in esecuzione, aggirando gli eventuali tempi di attesa dovuti al carico delle altre code di job a minor priorità per quel sito.

Sperimentato dapprima su un unico sito della Grid, il sistema di *code Emergency* ha garantito performance competitive per TEPHRA in termini di tempo d'esecuzione, ed è adesso in fase di implementazione sull'intera infrastruttura. Ciò è stato possibile grazie alla creazione di un gruppo di lavoro dedicato costituito dal personale dell'INGV-CT e Cometa presso INGV-CT e presso l'INFN, con risultati dei test molto soddisfacenti.

6. Risultati e conclusioni

Dai numerosi test effettuati, si è verificato l'effettivo funzionamento del modello TEPHRA sull'infrastruttura Grid, ricavandone una misura delle performance. Questo processo di *testing* ha permesso di ottimizzare le strategie di sottomissione per poter sfruttare al meglio la potenza dell'infrastruttura Grid, e di fare un raffronto con l'esecuzione dello stesso sul cluster locale.

La misura delle performance ha utilizzato come discriminante il calcolo dei tempi di sottomissione, prendendo come campione un job relativo ad una simulazione del modello TEPHRA. Sono stati presi in considerazione in maniera distinta i tempi dell'esecuzione dell'applicativo, ricavati tramite l'uso di appositi script in linguaggio *bash* fatti eseguire sul *master node* ad inizio e fine esecuzione, e i tempi globali dell'intero processo di sottomissione del job, ottenuti tramite le apposite funzioni del *Workload Management System*, la suite del middleware GLite per la gestione dei job sottomessi su griglia. Da questi, per differenza, si sono calcolati anche i tempi di "attesa", ovvero quelli che intercorrono tra la sottomissione del job e la fase di running vera e propria dell'applicazione, e che tiene conto dunque dei tempi in cui un job attende in coda dietro agli altri sottomessi su quel Computing Element.

Sono state effettuate diverse sottomissioni variando il numero di nodi, e sono state analizzate le medie dei tempi ottenuti. In Figura 8 è mostrato, mediante un grafico a linee, l'andamento medio dei tempi di esecuzione di una singola simulazione del modello TEPHRA al variare del numero di nodi di calcolo utilizzati.

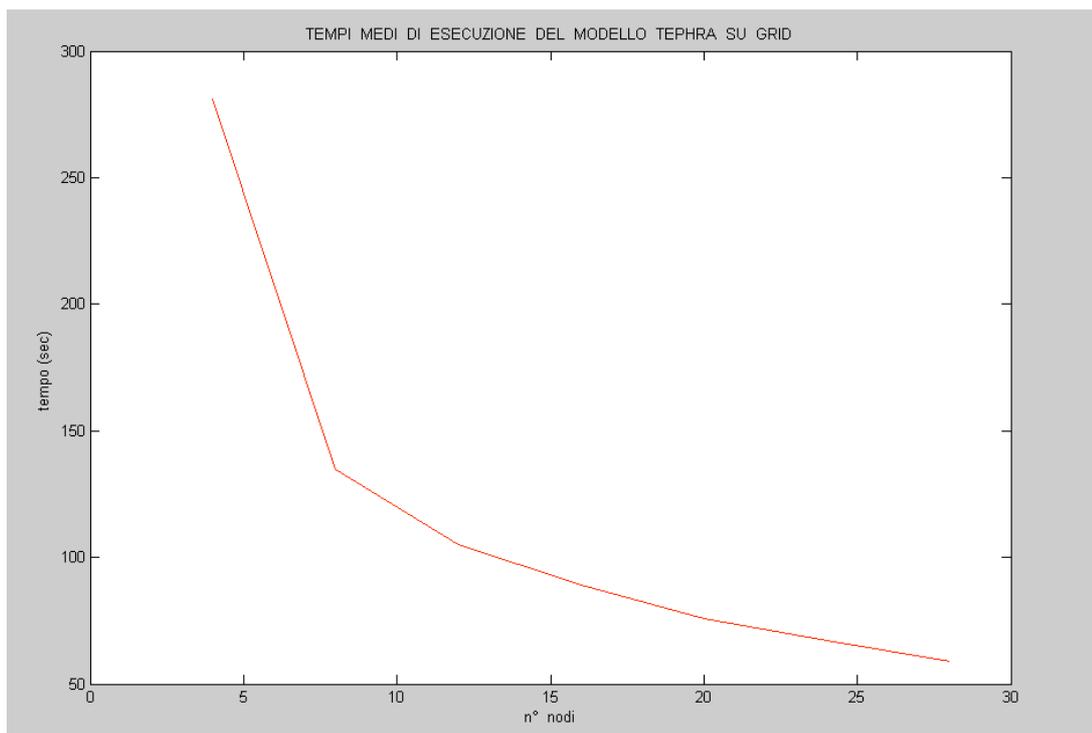


Figura 8. Tempi d'esecuzione di una simulazione del modello TEPHRA su Grid al variare del numero di nodi.

Come ci si attendeva i tempi di esecuzione sono risultati decrescenti all'aumentare del numero di nodi impiegati. Tuttavia il trend di decrescita diventa meno significativo superata la soglia dei 20 nodi.

Quindi è stata fatta una comparazione tra tali tempi ottenuti dai test su Grid e quelli ottenuti per la stessa simulazione sul cluster locale, variando allo stesso modo il numero di nodi utilizzati per il calcolo, così come illustrato in Figura 9. Dai confronti effettuati si è potuto riscontrare su Grid un vantaggio, in termini di tempo, del 26% rispetto a quanto raggiunto sul cluster.

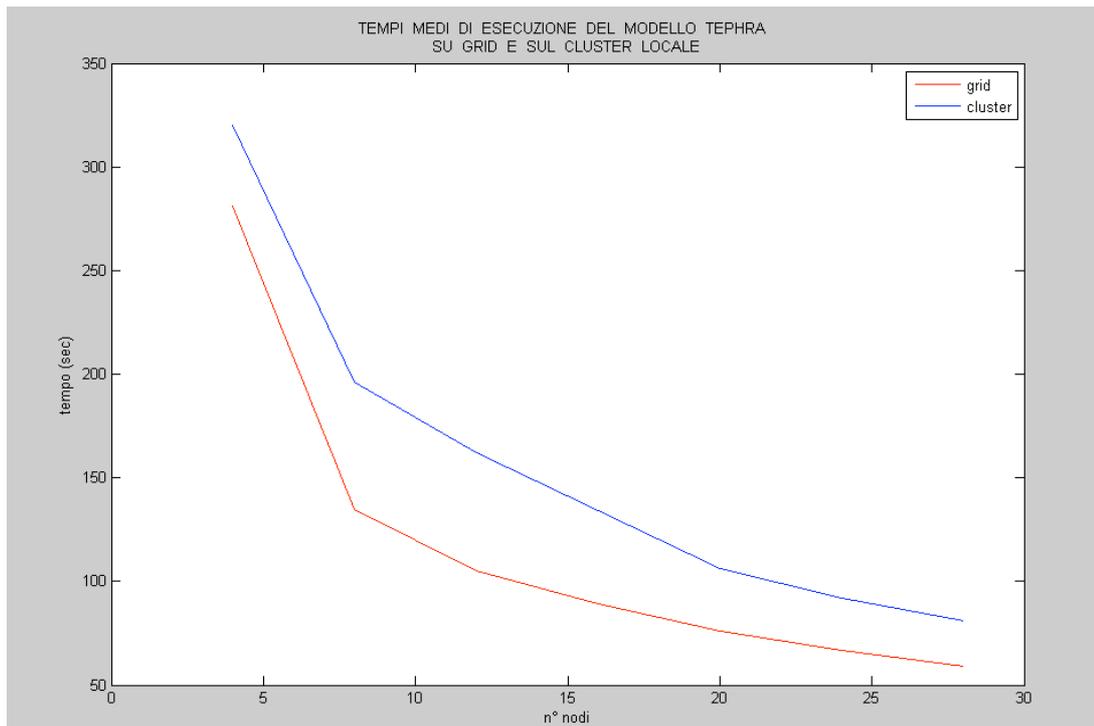


Figura 9. Tempi d'esecuzione a confronto su Grid (in rosso) e sul cluster locale (in blu) di una simulazione del modello TEPHRA.

Per sottomissioni regolari su Grid, senza particolari diritti di priorità, il vantaggio ottenuto sui tempi di esecuzione viene tuttavia vanificato se si vanno a considerare i tempi globali di sottomissione dei job, comprensivi dei tempi di coda. Infatti, mentre sul cluster, nel lasso di tempo in cui esso è riservato per l'esecuzione del modello TEPHRA, i tempi di attesa sono praticamente nulli, su Grid possono essere dell'ordine di diversi minuti. Essi tendono ad aumentare al crescere del numero di nodi, pur presentandosi comunque molto differenti tra loro (da pochi secondi a svariati minuti di attesa) e con caratteristiche di non predicibilità, in quanto dipendenti dal carico di lavoro del Computing Element in uso.

Così come illustrato dalla comparazione rappresentata nel grafico in Figura 10, l'uso delle *code "Emergency"* ha permesso di ridurre drasticamente i tempi di attesa riscontrati nelle comuni code, fino al 50% in alcuni casi.

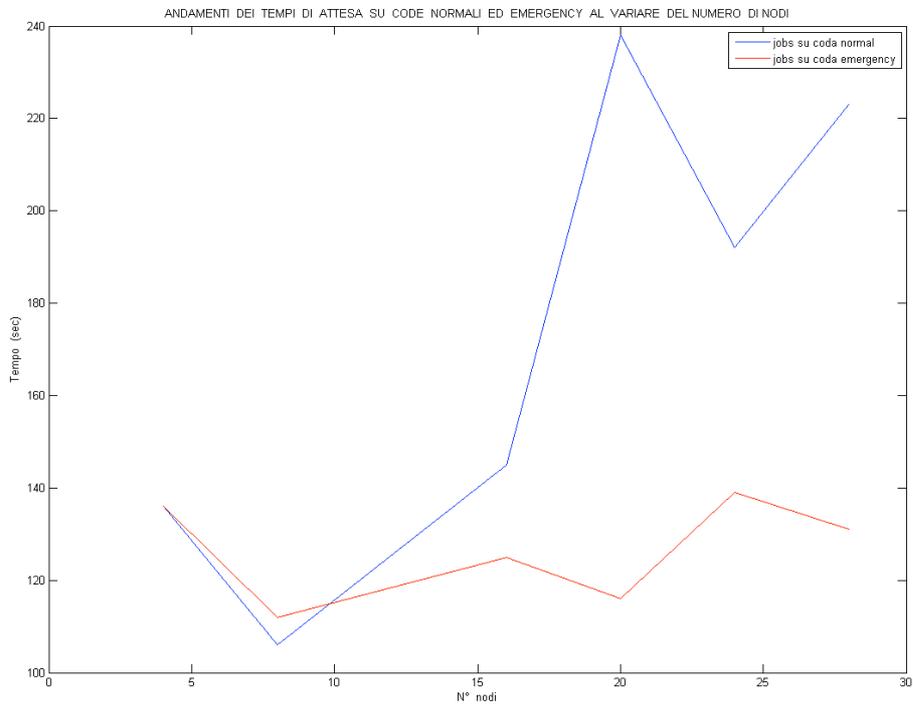


Figura 10. Confronto tra i tempi medi di *attesa* di un job su code normali e su code Emergency.

I tempi di attesa si sono stabilizzati, permettendo una maggiore predicibilità dei tempi complessivi, fattore indispensabile per la criticità temporale discussa sopra con cui ha a che fare l'esecuzione di TEPHRA presso INGV-CT, attestandosi attorno ad un range che va all'incirca dai 110 ai 140 secondi.

Comunque lo svantaggio, relativo alla presenza di questo tempo d'attesa, nell'utilizzare la Grid è compensato dal numero estremamente più alto di *core* di cui essa dispone, che permette, come spiegato sopra, di eseguire anche più simulazioni in contemporanea su job diversi lanciati simultaneamente.

Bibliografia

- Andronico, D., Del Carlo, P., Coltelli M. (1999), The 22 July 1998 fire fountain episode at Voragine Crater (Mt Etna, Italy), Volcanic and Magmatic Studies Group, Annual Meeting, Birmingham, 5 – 6 January.
- Andronico, D., Scollo, S., Cristaldi, A., Caruso, S. (2008), The 2002–03 Etna explosive activity: tephra dispersal and features of the deposit. *J. Geophys. Res.* doi:10.1029/2007JB005126.
- Barnard, S. T. (2004), Results of a reconnaissance trip to Mt. Etna, Italy: the effects of the 2002 eruption of Etna on the province of Catania, *Bulletin of New Zealand Society for earthquake engineering*, 37(2), 47-62.
- Barsotti, S., Neri, A., Scire, J. S. (2008). The VOL-CALPUFF model for atmospheric ash dispersal: 1. Approach and physical formulation, *J. Geophys. Res.*, 113 (B3), doi:10.1029/2006JB004623.
- Branca, S., Del Carlo, P. (2005). Types of eruptions of Etna volcano AD 1670–2003: implications for short-term eruptive behaviour, *Bull. Volcanol.*, 67, 732–742.
- Bonadonna, C., Connor, C.B., Houghton, B.F., Sahetapy-Engel, S., Hincks, T., Connor, L. (2005). Probabilistic modeling of tephra dispersion: hazard assessment of a multi-phase rhyolitic eruption at Tarawera, New Zealand. *J. Geophys. Res.* 110, B03203. doi:10.1029/2003JB002896.
- Casadevall, T.J. (1994). Volcanic ash and aviation safety. Proceedings of the First International Symposium on Volcanic Ash and Aviation Safety. U.S. Geological Survey Bulletin, vol. 2047, p. 450.
- Cioni, R., Longo, A., Macedonio, G., Santacroce, R., Sbrana, A., Sulpizio, R., Andronico, D. (2003), Assessing pyroclastic fall hazard through field data and numerical simulations: Example from Vesuvio, *J. of Geophys. Res.*, 108, 2063, doi: 10.1029/2001JB000642.
- Coltelli, M. (2006). Etna explosive eruptions and the INGV response to improve an ash-cloud monitoring system. Proceedings of Cities on Volcanoes 4, Quito, Ecuador, 22–27 January 2006.
- Costa, A., Macedonio, G., and Folch, A. (2006). A three dimensional Eulerian model for transport and deposition of volcanic ashes, *Earth Planet. Sc. Lett.*, 241, 634-647.
- D'Amours R. (1998). Modeling the ETEX plume dispersion with the Canadian emergency response model. *Atmospheric Environment* 32(24): 4335–4341.
- Draxler, R.R., Hess, G.D., 1998. An overview of the HYSPLIT- 4 modelling system for trajectories, dispersion and deposition. *Australian Meteorological Magazine*, 47, 295–308.
- Folch, A., Cavazzoni, C., Costa, A., Macedonio, G. (2008), An automatic procedure to forecast tephra fallout, *Journ. Volcanol. Geotherm. Res.*, 177, 767-777.
- Horwell, C.J., Baxter, P.J. (2006). The respiratory health hazards of volcanic ash: a review for volcanic risk mitigation. *Bull. Volcanol.* 69, 1–24.
- Macedonio, G., Pareschi, M.T., Santacroce, R (1988). A numerical simulation of the Plinian fall phase of 79 AD eruption of Vesuvius, *J. Geophys. Res.*, 93, 14817-14827.
- Macedonio, G., Costa, A, and Longo, V., (2005). A computer model for volcanic ash fallout and assessment of subsequent hazard, *Comput. Geosci-uk*, 31, 837-845.
- Rose, W.I., Bluth, G.J.S., Schneider, D.J., Ernst, G.G.J., Riley, C.M., Henderson, L.J., McGimsey, R.J. (2001). Observations of volcanic clouds in their first few days of atmospheric residence: the 1992 eruptions of Crater Peak, Mount Spurr Volcano, Alaska, *J. Geology*, 109, 677-694.

Ryall, D.B., Maryon, R.H. (1998). Validation of the UK Met. Office's NAME model against the ETEX dataset. *Atmospheric Environment*, 32(24), 4265–4276.

Sandu I, Bompay F, Stefan S. (2003). Validation of atmospheric dispersion models using ETEX data. *International Journal of Environment and Pollution* 19(4): 367–389.

Scollo S., Bonadonna C., Coltelli M., Del Carlo P. (2006). Probabilistic tephra fall-hazard assessment at Etna volcano, Italy. *Cities on Volcanoes 4*, 22-27 January 2006, Quito, Ecuador.

Scollo, S., Del Carlo, P., Coltelli, M. (2007). Tephra fallout of 2001 Etna flank eruption: analysis of the deposit and plume dispersion. *J. Volcanol. Geoth. Res.*, 160, 147–164.

Scollo, S., Tarantola, S., Bonadonna, C., Coltelli, M., Saltelli, A. (2008). Sensitivity analysis and uncertainty estimation for tephra dispersal models. *J. Geophys. Res.* doi:10.1029/2006JB004864.

Searcy, C., Dean, K., Stringer, W. (1998). PUFF: A high-resolution volcanic ash tracking model. *Journal of Volcanology and Geothermal Research* 80(1–2): 1–16.

Coordinamento editoriale e impaginazione

Centro Editoriale Nazionale | INGV

Progetto grafico e redazionale

Laboratorio Grafica e Immagini | INGV Roma

© 2009 INGV Istituto Nazionale di Geofisica e Vulcanologia

Via di Vigna Murata, 605

00143 Roma

Tel. +39 06518601 Fax +39 065041181

<http://www.ingv.it>



Istituto Nazionale di Geofisica e Vulcanologia