

Evoluzione dei Sistemi di Supercalcolo dal punto di vista della comunità Earth Sciences/Environment

Cineca, Bologna, 23.03.2018



Jost von Hardenberg
Istituto di Scienze dell'Atmosfera e del Clima, CNR

Principali settori di ricerca/progetti per risorse HPC

- Simulazioni con modelli numerici di clima globali o regionali
 - elevate risoluzioni
 - grandi ensemble di simulazioni (es. forecast stagionali/decadali)
 - lunghe integrazioni (es. scenari futuri, simulazioni storiche, paleoclima)
- Earth System modelling, biogeochemical and ecosystem dynamics
- Solid Earth modelling:
 - seismic wave propagation simulations
 - detection and localization of microseismicity with grid search methods
- Esempi di recenti progetti Europei:
 - H2020 PRIMAVERA (Hi-res simulations and Earth system modeling)
 - ERA4CS Medscope (seasonal predictions)
 - Progetto di Interesse NEXTDATA



Climate projects



- **CRESCENDO H2020 Project (2015-2020)**

- Climate scenario projections and experiments with an Earth-System Model version of EC-Earth

- **PRIMAVERA H2020 Project (2015-2020)**

- High resolution climate simulations with the EC-Earth model

- **ECOPOENTIAL (2015-2019)**

- Impact of environmental and climate change on mountain ecosystems



- **National Projects of Interest (PRIN) (2012-2015)**

1. Impacts of climate change on animal populations
2. Assessment of uncertainties in the hydrological cycle in Italy



- **NextDATA National Project of Interest (2011-2018)**

- Creation of a national archive for environmental and climate data



- **C3S-MAGIC Copernicus (2016-2018)**

- Climate service products for the analysis of global climate model data and in the water sector



- **ERA4CS MEDSCOPE (2017-2020)**

- Climate service products for seasonal to multiannual climate predictions in the Mediterranean

Sistemi di supercalcolo utilizzati/ criteri di accesso

Modalità di accesso:

Grant competitivi (PRACE, ISCRA, Special projects). Alcuni membri della comunità hanno convenzioni dirette con CINECA.

Macchine:

Cineca Marconi [Galileo]

Altre risorse europee (PRACE)

ECMWF cca (Special projects)



Supercomputing Projects



- **PRACE supercomputing project “Climate SPHINX” (2015-2016)**

- Climate simulations up to 16km resolution.
- 20M core hours (SuperMUC/LRZ).



- **PRACE project “Climate SPHINX reloaded” (2016-2017)**

- High resolution coupled climate simulations and role of stochastic physics
- 10M core hours (Marconi/CINECA)



- **Gauss EXPRESS project (2013-2014)**

- Hires climate simulations over Europe with WRF model
- 42 Mln core hours, 50 TB of output (LRZ/SuperMUC)



- **DATA SPHINX EUDAT Data Pilot project (2016-2018)**

- Creation of a widely accessible archive of climate model data and development of innovative additional services



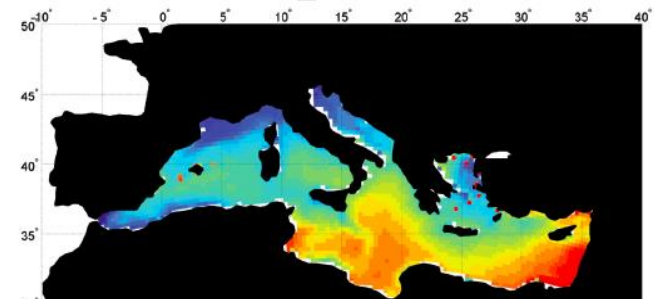
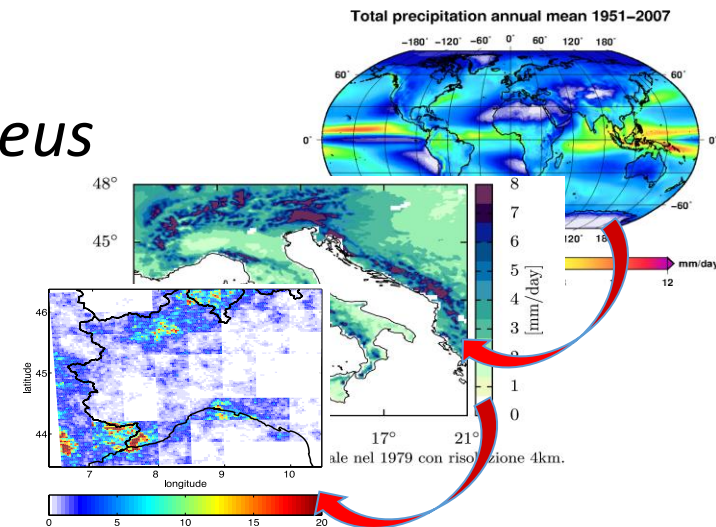
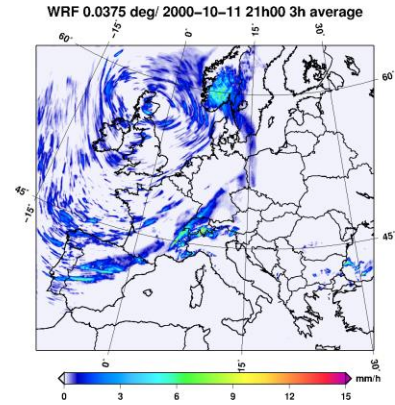
- **ECMWF special projects (2014-2016)-(2016-2018)**

- Impact of high resolution and stochastic physics in EC-Earth (75MSBU (=4.5M core hours) in 3 years 2016-2018)
- Sensitivity of multi-annual forecasts to model resolution (2014-2016) (57M SBU=3.5M core hours) in 3 years



Modelli ed applicativi

- **Modelli di clima globali ed Earth System**
EC-Earth, OpenIFS
- **Simulazioni oceaniche/biogeochimiche:**
NEMO, MedBFM (OGSTM-BFM-3DVarBio), MITgcm-BFM
- **Modelli clima e meteo regionali:**
RegCM4, WRF, Bolam/Moloch, Protheus
- **Solid Earth:**
SPECFEM3D Cartesian
(simulaz. onde sismiche in 3D con metodo agli elementi spettrali),
FPSM3D (simulaz. onde sismiche in 3D con metodo pseudospettrale di Fourier)



Servizi innovativi/necessità (1)

- Necessità di uso di risorse di calcolo tradizionali:
 - Disponibilità di nodi/tempi di coda
 - gestione degli input e output in preprocessing e postprocessing
 - tipicamente dell'ordine dei TB/utente: i processi di compressione/archiviazione necessitano di un numero limitato di cores performanti in I/O, disponibili senza attese. L'accesso parallelo al gss ottimizza questo tipo di operazioni.
- **BIG DATA:**
 - **postprocessing, archiviazione, trasferimento e distribuzione/pubblicazione** di centinaia di TB di risultati (P.es. simulazioni climatiche ad alta risoluzione - soprattutto legate a progetti internazionali quali CMIP6)
 - Disponibilità operativa di servizi quale un Nodo ESGF per la distribuzione dei dati

Servizi innovativi/necessità (2)

- Possibile interesse per **CLOUD HPC** per valutare se adatto agli scopi di ricerca
- **MACHINE/DEEP LEARNING** - interesse in prospettiva per:
 - estrarre informazioni da output di simulazioni numeriche di grande complessità (crescente, sia in termini di volume di dati multi-variati biogeochimici, sia di risoluzione spaziale).
 - Migliore capacità di analisi / post-processing rispetto ai paradigmi usati correntemente.
 - Es.: deep learning per il riconoscimento in tempo reale di micro-eventi dal flusso di dati di una rete di monitoraggio sismico

Modello di utilizzo:

Principalmente **batch** (pre-processing / run / post-processing)

Interesse per visualizzazione interattiva in-situ (es. tramite Paraview)

- per estrazione runtime di snapshot di variabili
- analisi interattive e visualizzazione di grandi dataset già prodotti

Modalità di accesso

- Necessità di una 'base' permanente per sviluppo, ricerca e gestione/analisi dati
- Convenzioni/partnership per sviluppo congiunto di soluzioni
- Interesse per modello cloud like:
 - ambienti di simulazione efficienti/replicabili
 - Tempi di attesa in coda brevi per sviluppo

Competenze specialistiche

- Esperti in programmazione parallela, software design
- Gestione, archiviazione e distribuzione grandi moli di dati (es. ESGF)
- Supporto professionale Cineca dedicato
- Progetti pilota comuni tra HPC e comunità scientifiche per sviluppo di soluzioni specifiche/identificazione di knowledge gaps e trasferimento di competenze/training. Esempi: *Pilot EUDAT, progetto SHARE, HPC-Tres*
- Necessità di dialogo continuo con esperti per capire come meglio usare le risorse disponibili (es. ottimizzazione code/richiesta risorse) e per distribuzione e trasferimento dati