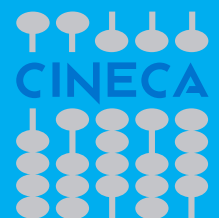
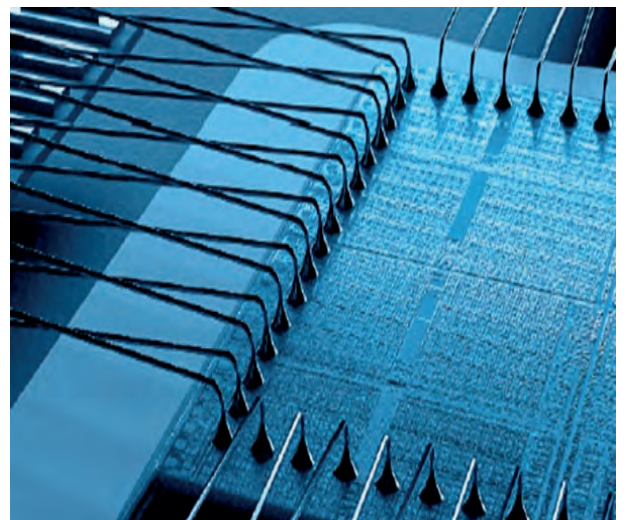
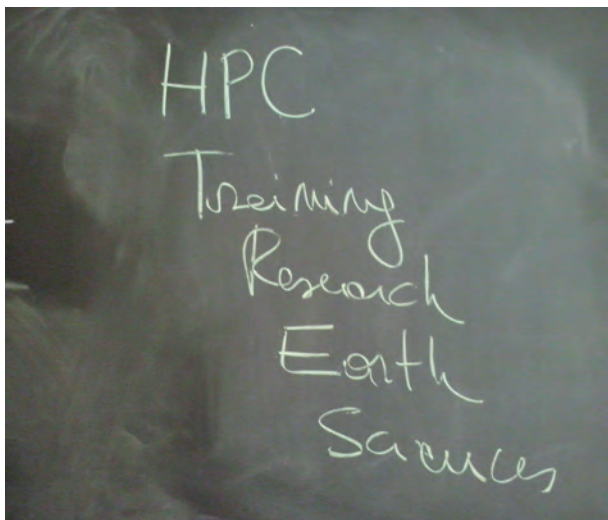


HPC TRES

High Performance Computing



Foreword	2
Topic A: HPC for Oceanography and Biogeochemistry modelling.....	3
Line A1. Biogeochemical forecast skill assessment for Operational Oceanography.....	3
Line A2. Evolution of Individual-Based Models towards exascale computing (“Exafish”)...3	
Line A3. Simulations of behaviours, responses, and adaptive and evolutionary characteristics of marine ecosystems.....	3
Line A4. Parallelization of hydrodynamic finite element models	4
Line A5. New algorithmic approaches for ocean models at exascale	4
Topic B: HPC for Solid Earth modelling.....	6
Line B1. Numerical simulation of earthquake ground motion in three-dimensional models of the Earth's crust.....	6
Line B2. Microseismicity data processing.....	6
Line B3. Seismic Laboratory set-up (SEISLAB)	7
Line B4. High level hybrid programming code for geodynamic simulations	7
Line B5. Multiphase flow modelling of explosive eruption dynamics.....	8
Line B6. Non-equilibrium dynamics in magmatic systems	8
Line B7. Environmental impact of explosive volcanic eruptions and volcanic hazard	9
Topic C: HPC for Earth System modelling.....	10
Line C1. High-resolution simulations of convective processes in the Earth System.....	10
Line C2. Optimization and parallelization of a new non-hydrostatic dynamical core for use in high resolution climate models.....	10
Line C3. Physical-Biogeochemical online coupling for Mediterranean Sea	10
Topic D: HPC for climate sciences	12
Line D1. Climate studies of natural drivers and anthropic pressures on Mediterranean Sea marine ecosystems	12
Line D2. High-resolution ensemble global climate simulations in the last 2000 years.....	12
Line D3. Coupling of marine bio-geochemical models to a regional Earth System model framework.....	13
Topic E: “BigData” in Earth Sciences.....	14
Line E1. Large scale data analysis for climate data.....	14
Line E2. Map-reduce approach to Regional Climate modelling pre/post-processing activities	14
Line E3. Data management, preservation and curation activities of the National Oceanographic Data Center (NODC)	15

Foreword

OGS (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - National Institute of Oceanography and Experimental Geophysics) and CINECA drew up an agreement¹ to implement a training program focussed on High Performance Computing (HPC) applications for Earth Sciences, the “**HPC Training and Research for Earth Sciences**” (**HPC-TRES**). The program is co-sponsored by the Minister of Education, University and Research (MIUR) under its extraordinary contribute for the Italian participation to activities related to the international infrastructure PRACE – The Partnership for Advanced Computing in Europe².

The major objectives of the program are: capacity building, enhancement of human resources, and advanced training in the fields of Earth System modelling (atmosphere, hydrosphere, lithosphere and biosphere) and numerical models, the latter considered as a strategic cross-cutting element for modelling activities. These objectives will be pursued through the use of national and European HPC infrastructures and services in the framework of PRACE, the optimization of algorithms and numerical codes, the optimal management of “Earth Sciences Big Data”, and the graphical visualization techniques for multidisciplinary applications in the Earth Sciences, also in the frame of the “Blue Growth” strategy.

Therefore, HPC-TRES agreement will establish, sponsor and oversee **training and research awards** (i.e. contributes for training and research activities set up according to current regulations that regulate scholarships – to be also used for masters and specialized courses in HPC -, research grants and PhDs) that will support the research lines **scientific plan³ of the HPC-TRES program**.

Some Italian research groups and institutions (INGV-Pisa, CNR/ISAC, CNR/IGG, CMCC, ICTP-ESP), where the external members of the HPC-TRES Steering Committee lead their research activities and personally contributed to define the scientific plan of the HPC-TRES program, have already subscribed the HPC-TRES initiative, offering logistic support and hospitality to the winners of the training and research awards sponsored by OGS.

OGS recognizes the scientific excellence of the hosting institutions as strategic to the educational and professional growth of the awarded young researchers. OGS strives to extend the participation in the scientific plan to other national research groups involved in HPC applications for Earth Sciences, in a way that these groups could propose new training and research activities in the next HPC-TRES Calls.

¹ OGS BoD document n. 7.4.2014, date 18/06/2014.

² www.prace-ri.eu

³ The scientific plan was written by the HPC-TRES Steering Committee, jointly appointed by OGS and CINECA (OGS BoD document n. 15.9.2014, date 18/12/2014) and composed by scientific experts involved in different HPC applications for Earth Sciences.

Topic A: HPC for Oceanography and Biogeochemistry modelling

Line A1. Biogeochemical forecast skill assessment for Operational Oceanography

In the frame of the European initiatives for Operational Oceanography (e.g. MyOcean, Copernicus Marine Environment Monitoring Services), this research line aims at providing numerical tools able to evaluate the reliability and the uncertainty of the operational forecasts produced by coupled physical-biogeochemical models. This activity has important consequences for: 1) support environmental agencies and other stakeholders with useful indications related to definition of the targets for the Good Environmental Status, as described in the Marine Strategy Framework Directive, and 2) assess the climate projections impacts on the marine biogeochemistry and ecosystems.

Biogeochemical models have different degrees of complexity, and are usually characterized by a high number of parameters: the global sensitivity analysis approach must be based on a large number of independent simulations, such as an ensemble of $O(10^4)$ members, in order to evaluate the model sensitivity to each parameter.

HPC can greatly support this activity by widening the ensemble set, initially to assess the model sensitivity (as in the PRACE GSENSMED project), and then to effectively support operational forecast. Post-processing tools constitute another

PI: Cosimo Solidoro (OGS) – csolidoro@ogs.trieste.it

Others: Paolo Lazzari, Giorgio Bolzon, Gianpiero Cossarini, Stefano Salon (OGS)

Line A2. Evolution of Individual-Based Models towards exascale computing (“Exafish”)

Currently most of the biogeochemical and ecosystem models are still based on the assumption that living organisms are homogeneously distributed over the volume defining a model domain cell, as if they were dissolved chemical compounds. In fact, while it is obvious that organisms are discrete entities and should be regarded as such, the use of Individual-Based Models in describing marine biogeochemistry and ecosystem dynamic has been hindered by the intrinsic complexity of the system to be described and the computational load required for comprehensive simulations over realistic spatial domains. The possibility offered by recent HPC developments might give new boost to this field. A challenge is to describe the time evolution of the interaction between different typologies of organisms within a food web, each type being defined by an ensemble of slightly different individuals. A second challenge is to use exascale computing to simulate the spatial-temporal dynamics of the resulting huge number of interacting organisms by means of a Lagrangian approach. A third challenge is the analysis of the huge amount of data provided as model output, possibly also in order to derive ecologically based subgrid parametrization of biogeochemical and ecological processes.

PI: Cosimo Solidoro (OGS) – csolidoro@ogs.trieste.it

Others: Paolo Lazzari, Donata Melaku Canu, Simone Libralato, Stefano Salon (OGS)

Line A3. Simulations of behaviours, responses, and adaptive and evolutionary characteristics of marine ecosystems

The goal of this research line is the development of new model paradigms aimed to simulate behaviours, responses, and adaptive/evolutionary characteristics (also referred to as “traits”)

of the marine ecosystems, with a strong dependency on natural drivers (e.g. increasing of the temperature field) and anthropic pressures (e.g. environment management). These new models will be able to provide more realistic estimates of the effects related both to global long-term climatic changes and to local environmental policies on marine ecosystems and their resources.

To this scope, innovative algorithms will be specifically designed to efficiently exploit the future heterogeneous HPC architectures, as for example a lattice-Boltzmann approach to solve transport equation on 3d meshes of $O(10^8)$ nodes. This activity will request new ad hoc parameterization/discretization developments and innovative programming techniques, binding together technological and scientific challenges and capacity building.

The added value will be constituted by the rise of a new ecological modelling community operating in an unexplored scientific field, able to define new evolutionary schemes never considered before. Therefore, we expect a significant commitment on the training and collaborative basis (e.g. the development of new expertise in computational ecology, the organization of a summer school on new paradigms in ecological modelling, international scientific collaboration).

PI: Cosimo Solidoro (OGS) – csolidoro@ogs.trieste.it

Others: Celia Laurent, Donata Canu (OGS)

Line A4. Parallelization of hydrodynamic finite element models

Hydrodynamic finite element models (FEMs) can be used to perform high-resolution simulations of areas characterized by complex geometry and bathymetry, such as lagoons, estuaries, coastal areas, lakes (though some FEMs are also applied for global ocean simulations⁴). Such models can be often legacy codes originally written for sequential execution, which appear today significantly limited given the possible applications able to exploit the larger availability of HPC resources.

An application of this research line is the parallelization of the Shallow water Hydrodynamic Finite Element Model (SHYFEM).

PI: Donata Melaku Canu (OGS) – dcanu@ogs.trieste.it

Others: Cosimo Solidoro, Leslie Aveytua, Eric Pascolo (OGS); Georg Umgiesser (CNR/ISMAR, Venezia)

Line A5. New algorithmic approaches for ocean models at exascale

Access to large computational capabilities will allow increasing both the resolution and the complexity of ocean models, exploiting resources capability. However, the scalability of the current models has to be improved. One of the main limits of the low scalability of legacy codes on many core architectures is the cost of data movement in terms of performance. Moreover, data movement is one of the main sources of energy consumption, so its reduction would benefit also energy-cost. The data movement issue can be intended at different levels, such as communication, memory access, I/O, etc. Its reduction needs not only the adoption of optimization techniques (i.e. communication and computation overlapping, search of the best trade-off between communication and computation, etc.). Indeed, the models have to be

⁴ e.g. FEOM, see

http://www.awi.de/en/research/research_divisions/climate_science/climate_dynamics/methods/finite_element_ocean_model_feom/

revisited from the algorithmic point of view, moving towards new algorithms, such as communication-avoiding and cache oblivious.

The research line aims at the re-design at algorithmic level of a target ocean model through the following steps: (a) the analysis of new algorithms reducing data movement and the evaluation of the benefits deriving from their adoption, using the mini-apps approach (as lighter versions of complex HPC applications, mini-apps are a flexible test bed to facilitate software development) (b) their integration into the target ocean model and (c) the performance comparison with the non-optimized code.

PI: Giovanni Aloisio (CMCC, Lecce) - giovanni.aloisio@cmcc.it

Others: Silvia Mocavero (CMCC, Lecce)

Topic B: HPC for Solid Earth modelling

Line B1. Numerical simulation of earthquake ground motion in three-dimensional models of the Earth's crust

Numerical simulations of earthquake scenarios based on the solution of the wave equation in three-dimensional viscoelastic heterogeneous media represent a promising tool for the studies of seismic hazard aimed at the mitigation of the seismic risk but they can be performed only with HPC resources.

In order to perform these simulations, our research group developed the parallel code FPSM3D (Klin et al. 2010), based on the method for the solution of differential equations known as "Fourier pseudospectral method". The accuracy of the method and the code was recently confirmed with the participation of our group to a comparative exercise of numerical simulations of ground motion generated by earthquakes in sedimentary basins (Chaljub et al., in press.). The parallelization of the code uses a hybrid scheme based on the two protocols MPI and OpenMP and is constantly updated with improvements in computational efficiency and in the complexity of the physical model to be treated. With the current implementation of FPSM3D on the supercomputer IBM BG/Q "Fermi" at CINECA a typical simulation requires a computational cost of the order of tens of thousands of core hours and several tens of simulations are needed in order to build up an earthquake scenario. Currently we plan an extensive application of numerical simulations for earthquake scenarios in the Po Valley (Italy).

Innovative versions of code suitable for simulating seismic wave propagation in stochastic media 3D are also under investigation.

PI: Peter Klin (OGS) – pklin@ogs.trieste.it

Others: Enrico Priolo (OGS)

Line B2. Microseismicity data processing

The goal of the project consists in the implementation and testing of a procedure that takes advantage of HPC resources for the automatic detection and location of microseismic events with a local network of receivers in a noisy environment. The use of automated procedures for the detection and location of microseismic events is essential, because of their high number. However microseismic signals are generally characterized by a very low signal to noise ratio that affects the performance of automated standard procedures. Several alternative approaches have been proposed in recent years, among which the most promising consist in the calculation of the probability distribution over a discretized parameter space. HPC techniques are necessary in order to apply efficiently these methods.

Our research group manages the seismic network of Collalto, an infrastructure aimed at monitoring the natural and induced seismicity in the field of natural gas storage called "Collalto Storage" near Treviso (<http://rete-collalto.crs.ogs.it>). With the methods under implementation we aim to alleviate the limitations due to the overlap of environmental noise on the recorded signals, to lower the magnitude of completeness and to allow a moment tensor inversion in order to make possible studies that discriminate between natural and induced seismicity. These studies are of increasing interest because of the growing perception of the risk associated with exploitation of earth resources.

PI: Peter Klin (OGS) – pklin@ogs.trieste.it

Others: Enrico Priolo (OGS)

Line B3. Seismic Laboratory set-up (SEISLAB)

This line of activity consists in setting-up a Seismic Laboratory (SEISLAB). The purpose of this project is to develop a logistic structure where OGS's researches and students from the University of Trieste can perform processing, interpretation, modelling and inversion of onshore and offshore seismic exploration data with an integrated software environment. Seislab aims to integrate commercial software such as Echos, Promax, KingDom Suite, Petrel, and others, with in-house codes developed by OGS's researches through a correct managing of the input/output file-format compatibility and the access to memory storage and external devices. The activity planned in 2015 consists in designing all the aspects concerning the laboratory (physical spaces, computers, software platform, personnel involved) and starting the practical research activity related to its implementation, e.g. porting of in-house codes for this new integration system.

Part of the codes considered for Seislab involve numerical modelling. The OGS, in fact, has a disposal a sequential code of 3D seismic modeling solving the viscoelastic, anisotropic wave equation with the free surface (air-ground interface). However, the run time required from this code is too large when realistic medium-large numerical meshes (several million points) are used. To overcome this limitation, the parallelization of the code is undoubtedly the most effective solution, and this activity constitutes a substantial part of the this line of research.

The work activity is articulated in different stages:

- Serial code optimization
- Code profiling to investigate the reasons that slow down the run of the code
- Development of the parallel code through programming paradigms such as MPI or OpenMP
- Debugging and validation of the performances of the code through speedup evaluation

All these steps must be performed on a HPC platform and with the aid of specific tools in both the phases of debugging and profiling.

PI: Davide Gei (OGS) – dgei@ogs.trieste.it

Others: Riccardo Geletti, Philippe Cance, Gianni Madrussani, Stefano Picotti, Josè Maria Carcione (OGS)

Line B4. High level hybrid programming code for geodynamic simulations

The deformation of Earth surface is the result of several forces that act inside the planet: complex models based on theoretic numerical simulations are needed to understand how the Earth surface evolves.

Starting from well-known numerical methodology already used among the geodynamic scientific community, a new software suite PyGmod (Python Geodynamic Modeling) has been developed in the last year. Unlike common simulation code written in Fortran or C the programs of this suite are pure Python for more than 90%, implementing a parallel schema that takes advantage of the new architecture of the last HPC machines. As the most recent HPC facilities are characterized by a distributed shared memory architecture, the code uses standard MPI coupled with a multi-threading parallelization algorithm to speed up some critical (atomic) parts that can benefit of it. Since the OpenMP A.P.I. is not directly accessible by Python, the threading system has been developed with Cython. A real time visualization

library, mainly based on Matplotlib, has been developed to inspect results during the computation.

The application simulates large-scale geodynamic processes involving the interaction of tectonic plates, the generation of magma, surface topography evolution, mineral phase changes, etc. The code will be mainly used to understand the dynamic of continental rifting.

Tests have been already done on multi-core machine in house and are now going live on some HPC facilities (GALILEO) provided by CINECA in the ISCRA PYGEO project.

PI: Nicola Creati (OGS) – ncreati@ogs.trieste.it

Others: Roberto Vidmar, Paolo Sterzai (OGS)

Line B5. Multiphase flow modelling of explosive eruption dynamics

This research activity focuses on the atmospheric dispersal of the products of explosive activity, namely volcanic gases and pyroclasts, i.e., dispersed particles produced by the fragmentation of the liquid magma in the volcanic conduit. To describe the interaction between the eruptive mixture constituents, this research has been primarily carried out by means of the three-dimensional Eulerian-Eulerian multiphase flow models (describing gas and individual particle classes as interpenetrating and interacting fluids) and equilibrium-Eulerian models, with potential new applications focusing on Eulerian-Lagrangian modelling. HPC research will be aimed at the understanding of specific multiphase flow processes whose features are still mostly unknown, and specifically: gas-particle turbulence and its control on the dynamics of volcanic plumes; the role of supersonic regimes and shock waves generated during explosive phenomena; the dynamics of particle and velocity stratification in pyroclastic density currents.

PI: Tomaso Esposti Ongaro (INGV-Pisa) - tomaso.espostiongaro@ingv.it

Others: Mattia de' Michieli Vitturi (INGV-Pisa)

Line B6. Non-equilibrium dynamics in magmatic systems

The different and unpredictable behaviour of volcanic eruptions is mostly related to the complex, non-linear dynamics of magma ascent along volcanic conduits. Non-equilibrium multiphase flow models based on the theory of thermodynamically compatible systems which uses a single set of equations to describe flow above and below the fragmentation level can describe the interaction between the liquid magma and the exsolved volatiles, potentially leading to the effusive/explosive transitions. Such kind of models is effective since it 1) is capable of treating both dilute and dense flow regimes; 2) allows multiple criteria for modelling the fragmentation transition; 3) quantifies the interaction between two phases (with two pressures and two velocities); 4) accounts for disequilibrium crystallization and degassing; and 5) allows open-system degassing in both vertical and lateral directions. The mathematical formulation and a numerical implementation based on a modular and object-oriented code (such as OpenFOAM), will allow to easily change the different constitutive terms – including, e.g., viscosity models, solubility models, crystallization models, drag terms, lateral degassing terms – to make extensive sensitivity analysis studies on the main eruption parameters.

Potential new applications will focus on numerical modelling of aggregation and breakage of a dispersed phase (bubbles in a liquid or particles dispersed in a gas) and numerical modelling of coupled fluid-rocks interaction in volcanic conduits.

PI: Tomaso Esposti Ongaro (INGV-Pisa) - tomaso.espostiongaro@ingv.it
Others: Mattia de' Michieli Vitturi (INGV-Pisa)

Line B7. Environmental impact of explosive volcanic eruptions and volcanic hazard

The assessment of explosive volcanic hazards (mainly associated to ash fallout from convective plumes and catastrophic pyroclastic density currents) requires a methodology to take into account the high level of uncertainty in initial and boundary conditions, the limited knowledge of the physics of volcanic phenomena and the limitations associated to numerical resolution. Coupling deterministic models (such as PDAC, Vol-CALPUFF and OpenFOAM) with statistical analysis tools (such as DAKOTA) has demonstrated to be effective for simple scalar models. However, coupling of complex 3D parallel models would require a careful code and workflow design, data assimilation, control of load balancing and appropriate sampling of the parameter space to reduce the computational cost of probabilistic analysis.

New HPC applications will focus on: inversion of observation data (such as thermal infrared and visible images of volcanic plumes); coupling between atmospheric and three-dimensional ash dispersal models; ensemble simulation of eruptive scenarios.

PI: Tomaso Esposti Ongaro (INGV-Pisa) - tomaso.espostiongaro@ingv.it
Others: Mattia de' Michieli Vitturi, Augusto Neri (INGV-Pisa)

Topic C: HPC for Earth System modelling

Line C1. High-resolution simulations of convective processes in the Earth System

Convection is an essential process in the Earth System, transporting heat and material in the atmosphere, the oceans, subsurface aquifers and the Earth mantle. At the same time, convective processes represent a paradigmatic example whose study has led to the development of many new ideas on the nonlinear dynamics of complex systems. The simulation and analysis of convective processes requires sophisticated simulation methods and powerful computing resources, to resolve the many scales at work. This research project is devoted to the development and implementation of numerical methods for high-resolution simulation of convective processes in the Earth System.

Research activities will include a first part devoted to simulating convection in idealized configurations, and a subsequent phase dealing with the application to a specific geophysical setting (atmosphere, ocean, or subsurface flows), to be defined based on the background of the fellowship recipient and the research priorities. From a technical viewpoint, this line aims to contribute to the development and improvement of numerical codes for convection simulations. From a scientific viewpoint, the research will lead to an increased understanding of convective processes in the Earth System and to specific applications.

PI: Antonello Provenzale (CNR/IGG, Pisa, Torino) - antonello.provenzale@cnr.it
Others: Marco Doveri (CNR/IGG, Pisa)

Line C2. Optimization and parallelization of a new non-hydrostatic dynamical core for use in high resolution climate models

The next frontier in regional climate modeling is the development of models that can reach the convection permitting horizontal resolution of 1-5 km and can thus provide climate information at scales more relevant for impact applications. This requires the use of efficient and scalable non-hydrostatic dynamical cores in climate models so that long climate runs at these resolutions can be conducted. Towards this goal, a new non-hydrostatic dynamical core is being developed at ICTP, based on a finite element, semi-lagrangian, semi-implicit numerical framework, which is highly accurate without having stability constraints. A key aspect of this development is the capability of this code to scale efficiently on massively parallel machines. The aim of this project is to optimize the code of this dynamical core in order to make it highly scalable on different HPC platforms and thus applicable in a wide climate modeling context, including the new version of the ICTP regional model RegCM5.

PI: Filippo Giorgi (ICTP) - giorgi@ictp.it
Others: Graziano Giuliani (ICTP)

Line C3. Physical-Biogeochemical online coupling for Mediterranean Sea

One of the main HPC challenges in oceanography modelling is the increase of the model spatial resolution to potentially allow more realistic simulations of the coastal areas. In an exascale paradigm, a multi-model approach will be able to simultaneously simulate the Mediterranean Sea at different spatial resolutions and on different sub-regions, conveniently calibrating the resolution increment and the boundary conditions between the basin and sub-

basin scales. Online coupling between physical and biogeochemical components is envisaged as one of the issues to be tackled

The goal of this research line is to start developing an online coupling between biogeochemical and physical models for the Mediterranean Sea to be applied both for operational applications and climate projections. Background experience for the present activity is the online coupling between MITgcm and BFM implemented at OGS for the Northern Adriatic Sea with collaboration of UTMEA-CLIM section of ENEA.

The perspective view is the development of a regional system for operational forecasts specifically designed for the Northern Adriatic Sea, including numerical models (coupled model physics-biogeochemistry, eventually extended to upper trophic levels) and observational data sets necessary for data assimilation (of surface chlorophyll, in-situ data, surface currents).

PI: Gianpiero Cossarini (OGS) – gcossarini@ogs.trieste.it

Others: Stefano Querin, Valeria Di Biagio, Cosimo Solidoro (OGS); Gianmaria Sannino (ENEA)

Topic D: HPC for climate sciences

Line D1. Climate studies of natural drivers and anthropic pressures on Mediterranean Sea marine ecosystems

Mediterranean Sea is an extremely heterogeneous sea basin in terms of physical forcings and biodiversity, and can be considered as a very sensitive area to the climate change effects on its ecosystems. The projections of the socio-economical effects related to climate change can be provided by integrated regional modelling (or Regional Earth System modelling), based on the coupling between ocean-atmosphere models, biogeochemistry models that involve the whole trophic chain, from plankton to big predators (“End2End” models), and socio-economical models. This approach, partially implemented in some EU-projects (e.g. PERSEUS, OPEC), is characterized by the computational challenge that could efficiently support simultaneously interacting, inter-dependent codes running on high-resolution grids for multi-annual integration periods, and that produce large amount of data.

Given the high uncertainty related to the sensitivity of the biogeochemical models to their parameters, the aim of this research line is the exploration of an ensemble of multi-decadal simulations (generated according to different initial and boundary conditions) of the Mediterranean Sea biogeochemistry, in order to evaluate the impact of the natural drivers (linked to climate change) and anthropic pressures (linked to environmental policies) on the marine ecosystems. This topic is extremely strategic in the frame of the Marine Strategy Framework Directive (MSFD)⁵.

PI: Stefano Salon (OGS) – ssalon@ogs.trieste.it

Others: Paolo Lazzari, Giorgio Bolzon, Simone Libralato, Donata Melaku Canu, Cosimo Solidoro (OGS)

Line D2. High-resolution ensemble global climate simulations in the last 2000 years

The reconstruction of past climatic fluctuations is an essential step to obtain the baseline of climate variability against which to project current climate change. Several proxy records are now available for the last few Millennia, with a specific focus on the past 2000 yr, as for example indicated by the international PAGES-2k initiative. Parallel to the proxy data reconstructions, simulations of climate variability in the last Millennia are necessary. Technically, however, such simulations - if performed at high spatial resolution- require HPC approaches, since one single simulation is not sufficient and an ensemble of possible climatic histories is usually needed. This research project focuses on the implementation, tuning and use of the high-resolution global climate model EC-Earth for ensemble paleoclimate simulations. The goal is to obtain an ensemble of high-resolution (0.75°) simulations of global climate in the last 2000 years with different solar and land-use forcing reconstructions, to obtain estimates of natural climate variability in the centuries before anthropogenic greenhouse gas emissions.

PI: Jost von Hardenberg (CNR/ISAC, Torino) - j.vonhardenberg@isac.cnr.it

Others: Antonello Provenzale (CNR/IGG, Pisa, Torino)

⁵ http://ec.europa.eu/environment/water/marine/directive_en.htm

Line D3. Coupling of marine bio-geochemical models to a regional Earth System model framework

The interest is steadily growing within the regional climate modelling community on the development of fully coupled Regional Earth System Models (RESMs), aimed at addressing climate system interactions at the regional to local scale. Within this context, a component noticeably missing in most current RESM efforts is marine biogeochemistry, which on the other hand is of high interest in global change studies.

In particular, ICTP is currently developing a coupled RESM including atmosphere (RegCM), ocean (MIT-GCM) and land (CLM) components. At the same time, OGS has developed and used a range of marine biogeochemistry and ecosystem models, so that a clear opportunity exists to develop a collaboration aimed at the integration of a marine biogeochemistry component within the ICTP regional model system.

Target of this project is to carry out such an exercise with specific applications to the Mediterranean region, addressing in particular the issue of the sensitivity of coupled processes to the complexity of the biogeochemistry model components.

PI: Filippo Giorgi (ICTP) – giorgi@ictp.it

Others: Cosimo Solidoro (OGS)

Topic E: “BigData” in Earth Sciences

Line E1. Large scale data analysis for climate data

In several eScience domains the analysis and mining of large volumes of data is increasingly becoming central to scientific discovery. In particular, the multidimensionality, volume, complexity, and variety of scientific data require specific solutions able to (1) support the management and analysis of large datasets, (2) provide array-based functionalities, and (3) support the scientific analysis process through parallel solutions able to deliver results in (near) real-time.

The main objective of this research project is to identify, extend and apply scalable big data analysis solutions running data intensive tasks on High Performance Computing (HPC) machines to address scientific analysis needs and issues in (near) real-time. The work will relate to (1) requirements gathering concerning the analysis of large volumes of scientific data (e.g. climate/weather), with a specific focus on multidimensional datasets; (2) the analysis and identification of big data platforms suitable for scientific data analysis and visualization to perform a cross-domain evaluation; (3) the extension and customization of the selected big data platforms for climate datasets, to support a large number of scientific use cases; specifically, these platforms will address scalability in terms of new storage models and data partitioning, efficiency in terms of parallel solutions and interoperability requirements by adopting well-known standards at different levels (data formats, communication protocols, etc.).

PI: Giovanni Aloisio (CMCC, Lecce) - giovanni.aloisio@cmcc.it

Others: Sandro Fiore (CMCC, Lecce)

Line E2. Map-reduce approach to Regional Climate modelling pre/post-processing activities

This proposal refers to the idea of evaluating some of Big Data tools/techniques developed for big data activities for industry, in the scientific area of Earth Science. More specifically the goal of this proposal is to test and investigate the use of Hadoop, a widely used Big Data framework, its Distributed File System (HDFS) and its MapReduce data-driven computational engine to run the preprocessing and the post-processing phase of a climate experiment involving a regional climate modeling tool on a HPC cluster.

In the context of Climate modeling, a considerable amount of time is spent in processing huge volume of data, either in preparing model input and analyzing the large amount of data produced by the simulations. Application codes used to process the data are usually written in Fortran language, and most of the datasets used in the climate modeling community are in the NetCDF format making the adoption of Hadoop not straightforward at all [1].

To achieve the overall goal Hadoop’s logic should be modified to allow in a transparent way both 1) code execution using all the pre/post processing tools and 2) dataset management on HDFS without any modification on the original dataset. To standardize the execution environment on Hadoop, and to make the approach as much portable as possible, the usage of Container-based Virtualization with Docker will also be considered.

The proposed case study will focus on the RegCM Regional Climate Model tool developed by the ICTP and its preprocessing and post-processing programs, but the success of this project would bring benefits also to the entire Earth Science community when dealing with similar use cases, and to the wider extent of filling the gap between modern Big Data platforms and scientific codes in general.

PI: Graziano Giuliani (ICTP) - ggiulian@ictp.it
Others: Stefano Cozzini (CNR/IOM, Trieste); Filippo Giorgi (ICTP)

Line E3. Data management, preservation and curation activities of the National Oceanographic Data Center (NODC)

EMODNet Chemistry (coordinated by OGS) is the European Marine Observation and Data network focused on the groups of chemicals relevant for MSFD Descriptors 5, 8 and 9, aiming to unlock and make available European marine data resources from diverse sources in a uniform way.

The assembled data, data products and metadata are hosted on a structured repository on a cloud environment, made available by CINECA. A series of nodes presently host horizontal gridded maps produced with DIVA (Data-Inverse Variational Analysis) made available through an in-house WMS service and in-situ data, properly discovered, analyzed and visualized with a standard, OGC compliant, WPS service. The data discovery and access service enable users to have a detailed insight of the availability and geographical extent of archived data. This repository hosted in the CINECA cloud represents a wealth of information scientifically aggregated and validated, accessible via the EMODnet Chemistry portal.

The Italian National Oceanographic Data Center (NODC) managed by OGS is going to be replicated at CINECA, together with a web service environment to satisfy the needs of European projects asking for continuous access to data and metadata and national collaborations.

PI: Alessandra Giorgetti (OGS) – agiorgetti@ogs.trieste.it
Others: Alberto Brosich, Alessandro Altenburger, Matteo Vinci (OGS)
Links with: Maris, Deltares, University of Liege