Models coupling with specific focus on ocean physics and biogeochemistry

Paolo Lazzari
plazzari@ogs.trieste.it

OUTLINE

- ✓ Model coupling
- ✓ What is biogeochemical modelling
- ✓ The OFFLINE coupling
- ✓ Some technical issues

The need for model coupling

http://www.esa.int/spaceinimages/Images/2006/09/The_Changing_Earth

- Atmosphere (air)
- Hydrosphere (groundwater, glaciers, oceans, lakes and streams)
- Cryosphere (sea ice, glaciers, snow)
- Biosphere (sum of all ecosystems, presence of DNA)
- Pedosphere (soil)
- Lithosphere (crust and uppermost solid mantle)
- Anthroposphere (human activities)

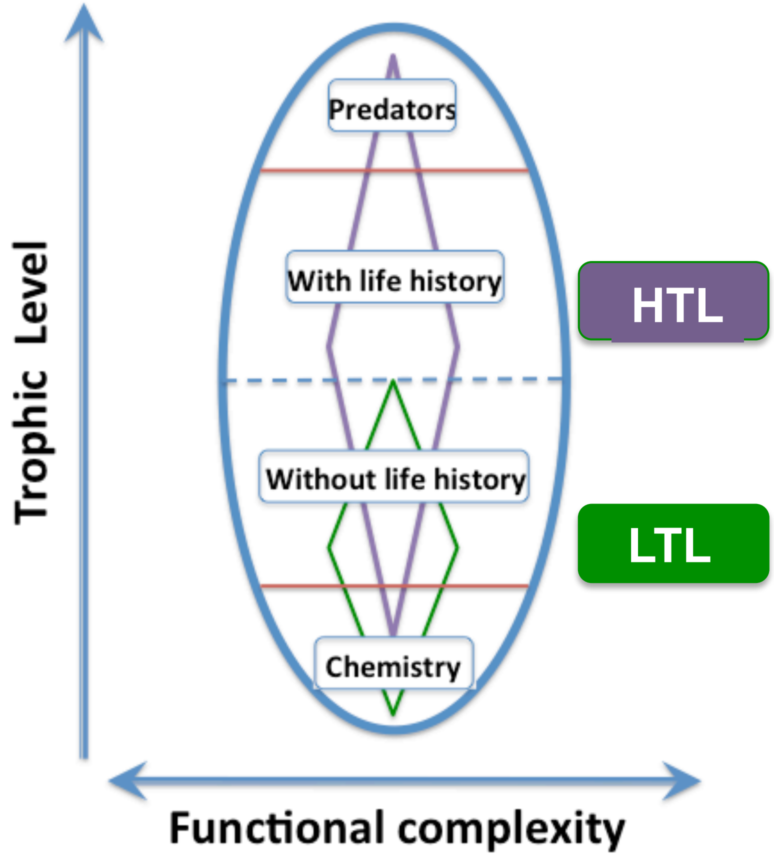
"**COUPLING**" → flow of mass, momentum, energy between spheres

… In the marine biosphere itself …

Many modules/approaches are present, each related to different scientific communities

- Plankton organisms (Lower Trophic Levels)
- Fishes (Higher Trophic Levels)
- Benthic/Sediment models ("the pedosphere @ the ocean bottom")

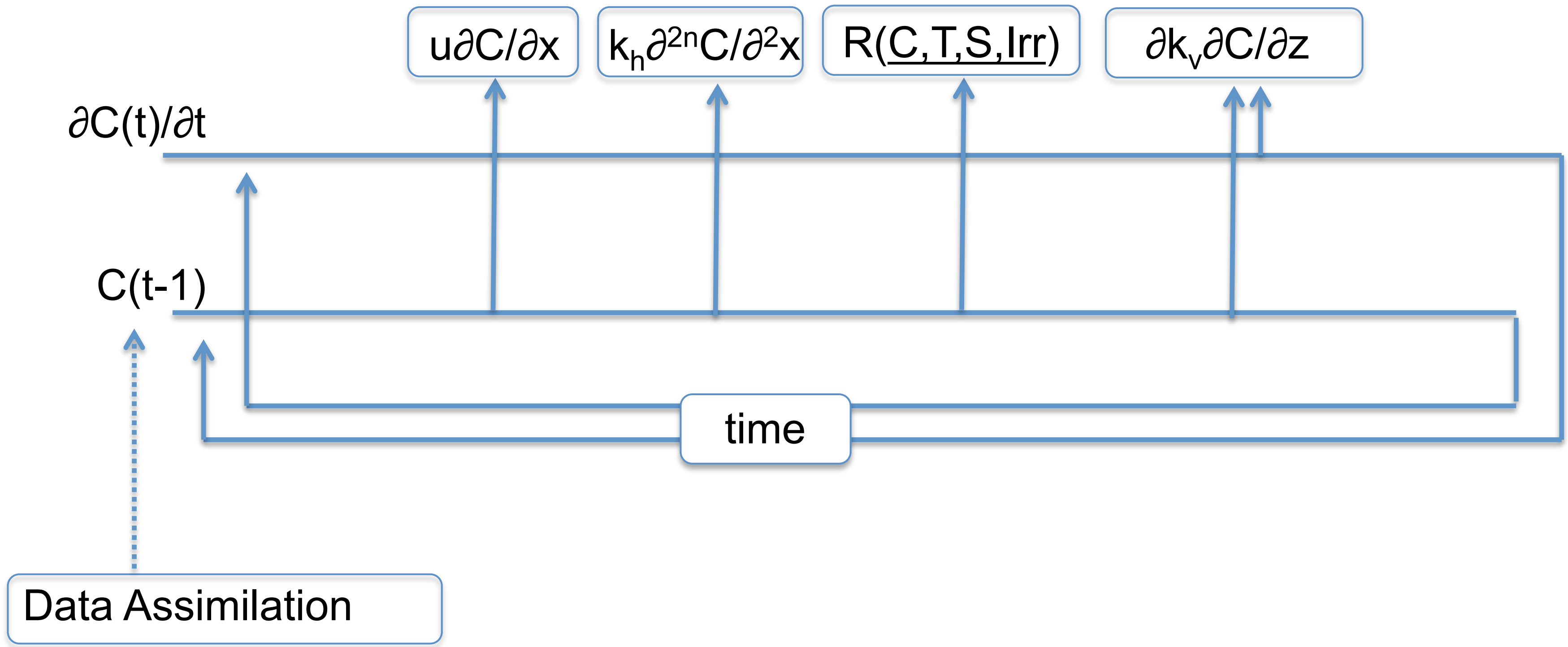
"**COUPLING**" → flow of mass due to predation dynamics or sedimentation/resuspension



De Young, 2004

… Coupling in terms of forcings …

Forcings coming from other models (e.g. circulation in a transport model), or model trajectory correction (e.g. 3DVAR)



“**COUPLING**” →forcing from an external model on a specific operator e.g. advection

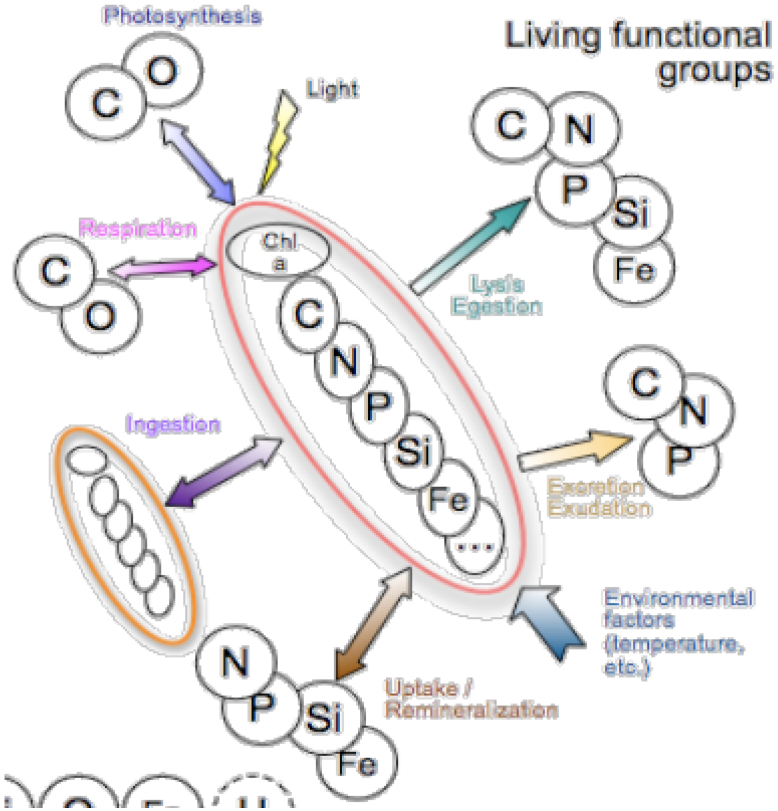
What is marine biogeochemistry?

This talk → focus on marine biogeochemistry

It is about the distributions of chemical concentrations in the Sea and the processes that control them

What is the mean abundance of key elements

What is the spatial and temporal variance of such distributions, what control such variability?



Vichi et al., 2007

Spatial scales: from global ocean to coastal areas
Temporal scales: hindcast mode, forecast, scenario projection

Basic equations

Eulerian framework, conservation equations for chemical concentrations ***C***

$$\frac{\partial C}{\partial t} = \left.\frac{\partial C}{\partial t}\right|_{advection} + \left.\frac{\partial C}{\partial t}\right|_{diffusion} + R(C, T, S, Irr, \dots)$$

R → reaction terms (Source minus Sink)

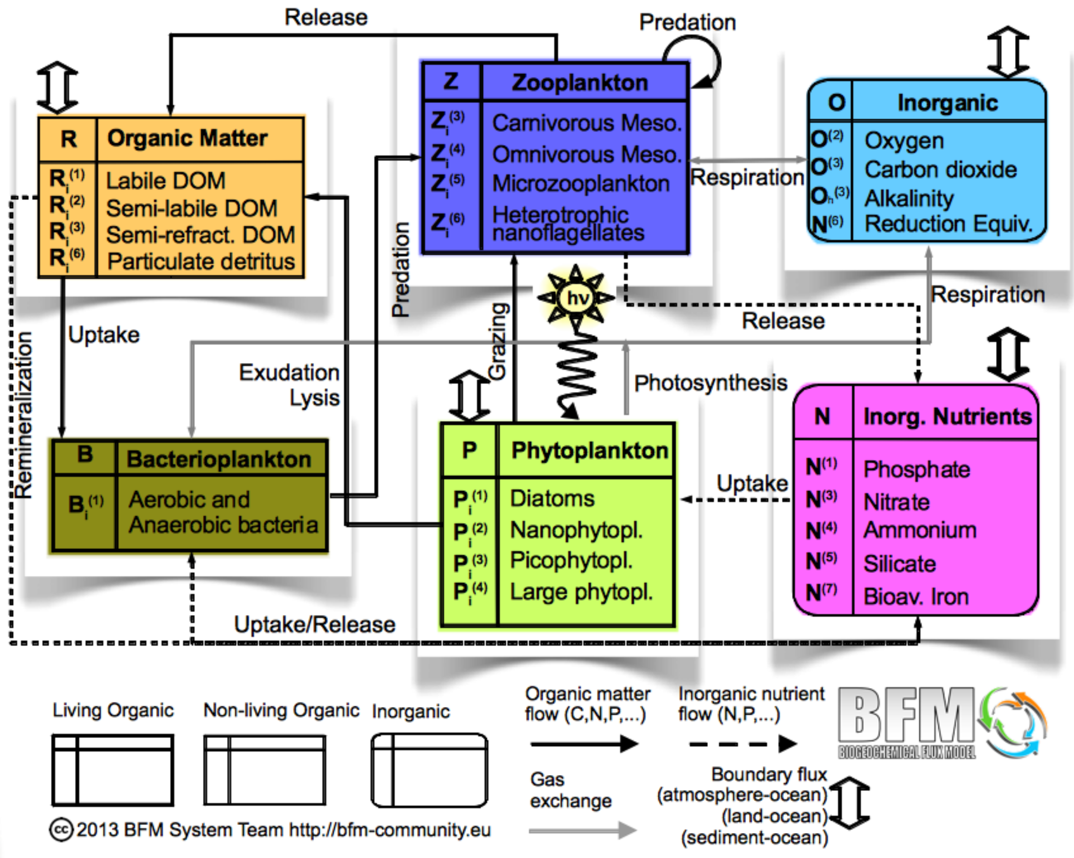
$$\frac{\partial C}{\partial t} = -\mathbf{U} \cdot \nabla C + \nabla(\mathbf{D} \nabla C) + R(\mathbf{C}, T, S, Irr, \dots)$$

"**COUPLING**" →OGCM
(Navier Stokes equation solver)
es. NEMO
MITgcm
ROMS

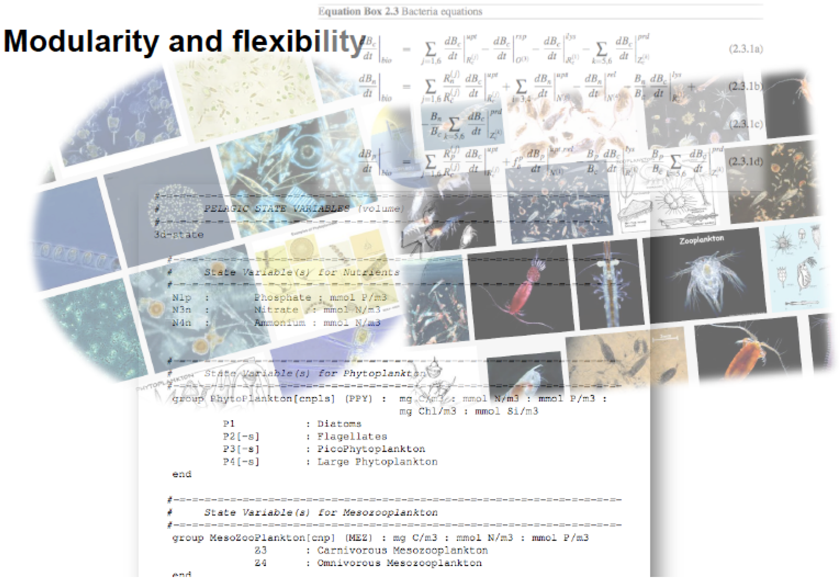
"**COUPLING**" →REACTOR
(Biogeochemical fluxes)
es. BFM
ERSEM
…

Biogeochemical Reactors

Biogeochemical modelling is usually a problem oriented discipline many different models with different parameterization /configurations



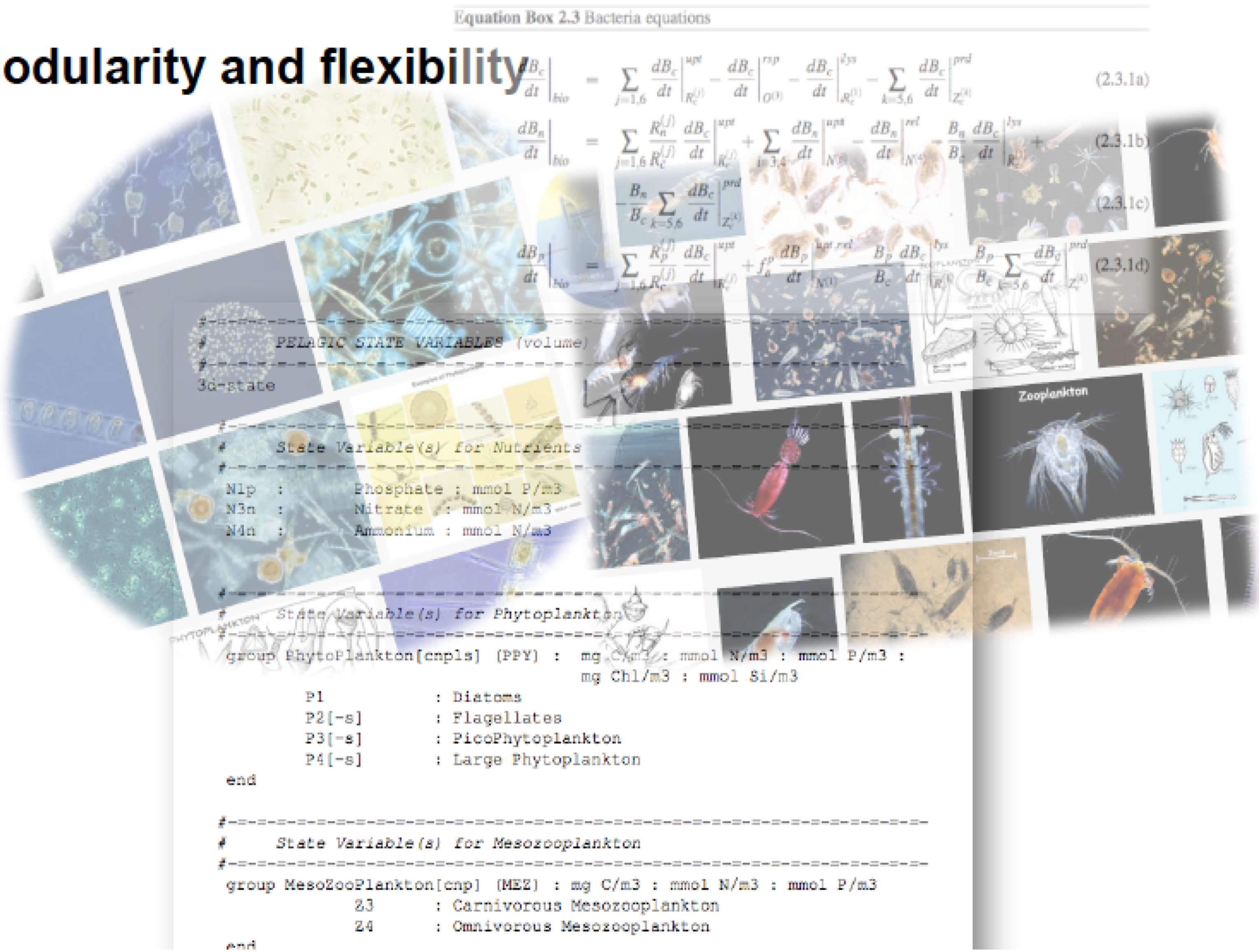
Modularity and flexibility



The reactor shoud be modular by itself to be adapted to specific problems

Biogeochemical Reactors

Modularity and flexibility



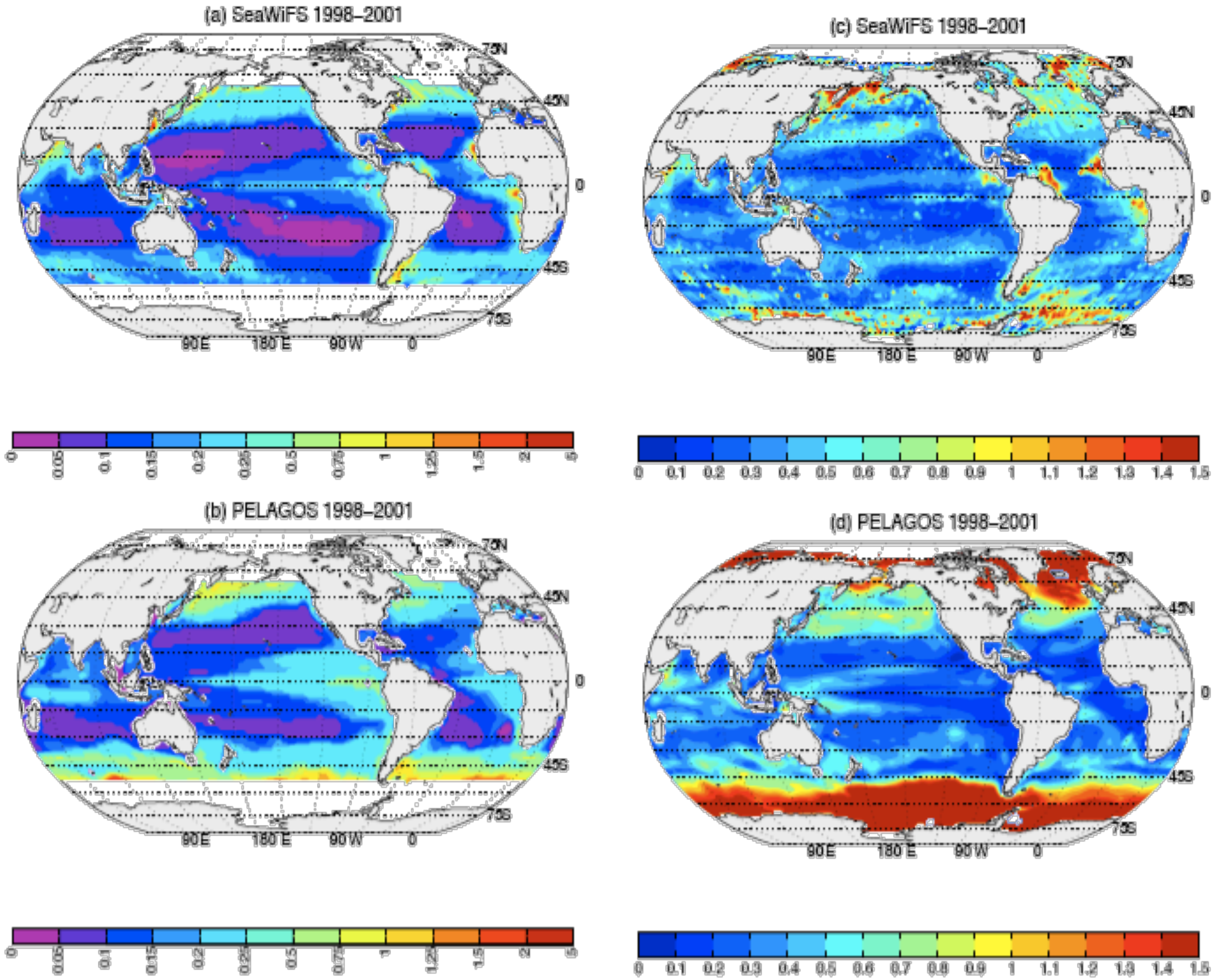
… Some examples ...

Global Ocean applications

- ✓ BFM coupling with the GCM NEMO at 0.25° and 2°
- ✓ Hindcast simulations of the global ocean biogeochemistry (Vichi et al. 2007a,b; Vichi and Masina, 2009)
- ✓ Biogeochemical cycles in the Earth System under current and future climate conditions with the CMCC Earth System Model (Vichi et al., 2011; Patara et al., 2011; Patara et al., 2012)

PELAGOS

(PELAgic biogeochemistry for Global Ocean Simulations)

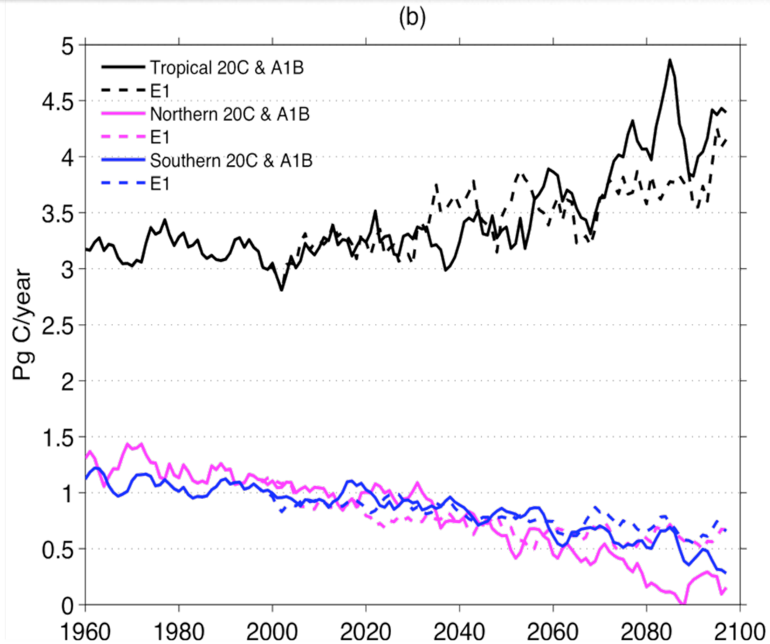


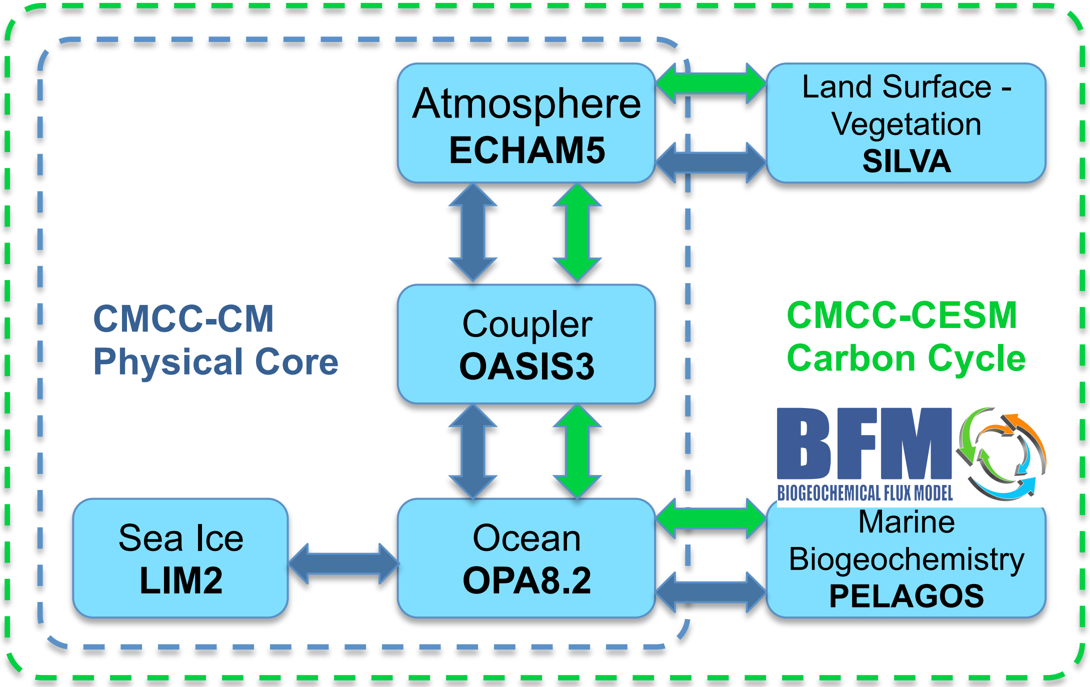
cmcc Centro euro-Mediterraneo sui Cambiamenti Climatici

OGS

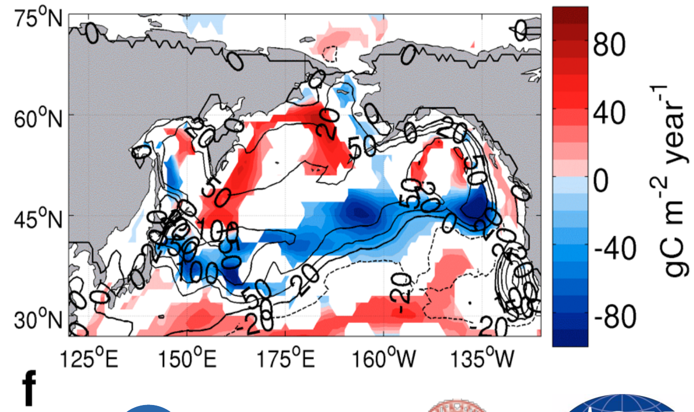
Earth System Modelling

PELAGOS is the marine biogeochemistry component of the CMCC-CESM Carbon Earth System Model
that participated to the Climate Model Intercomparison Project Phase 5 (Cagnazzo et al., 2013)

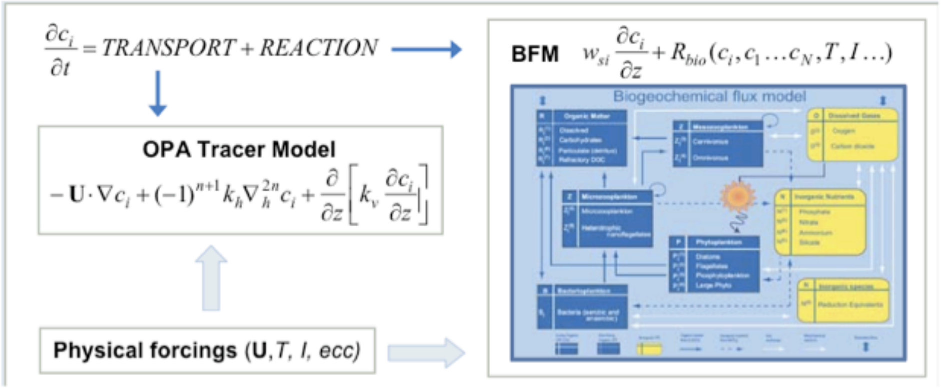




Scenario changes of Net Community Production in the Pacific



BFM in the Mediterranean Med



- ✓ Horiz. Res. = 1/8°
- ✓ Vert. Res. = 43/72 levels
- ✓ Time Res. = 1800 s
- ✓ 1 year simulated in 2 hours

Offline approach:

coupling with precomputed physical fields from OGCM

Online approach:

Runtime coupling with OGCM (e.g. NEMO)

cmcc Centro euro-Mediterraneo sui Cambiamenti Climatici

Model description and forcings

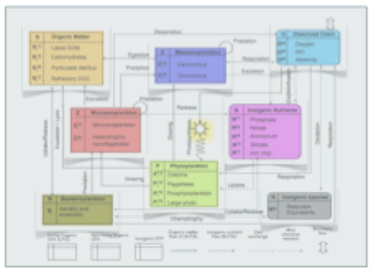
CIRCULATION MODEL

-OGCM: NEMO (OPA 9) (Madec 2008 http:/www.nemo-ocean.eu Oddo et al. 2009)
-Horizontal Resolution: 1/16 deg ~ 7 Km
-Vertical Resolution: Z-coordinates, 71 levels (partial steps)
-Free run: no relaxation to climatology, full freshwater flux (major rivers), no heat flux correction
-Parallel simulations (on-line)

BIOGEOCHEMICAL FLUX MODEL

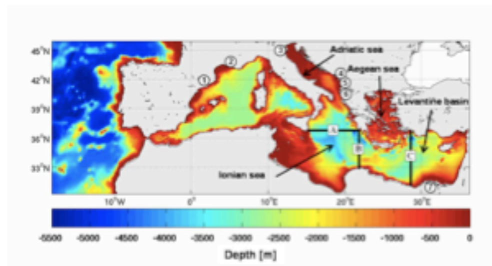
-BFM: Biological Flux Model (Vichi et al. a,b 2007)
-Carbon based multi-nutrient food web description
-Carbon, Nitrogen, Phosphorus and Silica cycles
-Potential for multiple nutrient co-limitation (Nitrogen, Phosphorus and Silica)

- phytoplancton
- mesozooplancton
- microzooplancton
- bacteria



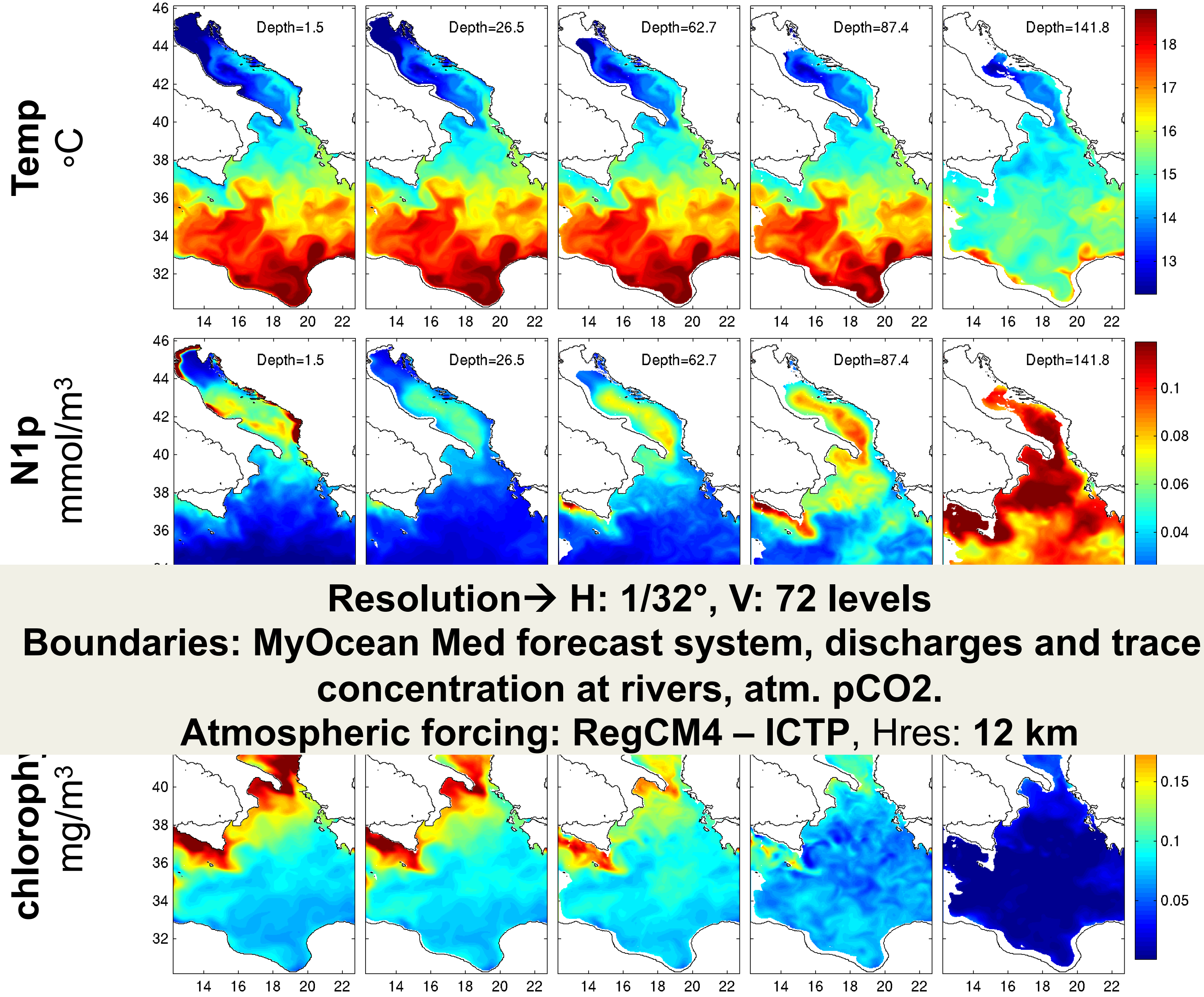
SET UP

-Physical model settings: ECMWF ERA40 atmospheric forcing functions
-Initial conditions for nutrients and oxygen: annual OA climatologies from SEADATANET project (http://www.seadatanet.org) merged with World Ocean Atlas climatology in the Atlantic box
-Initial conditions for biology: homogeneous guesstimates with vertically-distributed analytical profiles
-Nutrient River input: data from Sesame Project (Ludwig et al. 2009)



Test case: Adriatic sea + Ionian Sea

Winter conditions:



Resolution→ H: 1/32°, V: 72 levels
Boundaries: MyOcean Med forecast system, discharges and tracers concentration at rivers, atm. pCO2.
Atmospheric forcing: RegCM4 – ICTP, Hres: **12 km**

BFM BIOGEOCHEMICAL FLUX MODEL

… Different approaches …

INTERNAL COUPLING

Elements are in the **same** program unit

ON-LINE
Data exchange through working memory (es .mod)

DIRECT

EXTERNAL COUPLING

Elements are in **different** program units
e.g. both can be run indipendently

OFF-LINE
Date exchange through external storage (files)
Usually one way

ON-LINE
Date exchange through working memory

DIRECT

INDIRECT
Use of a **coupler**

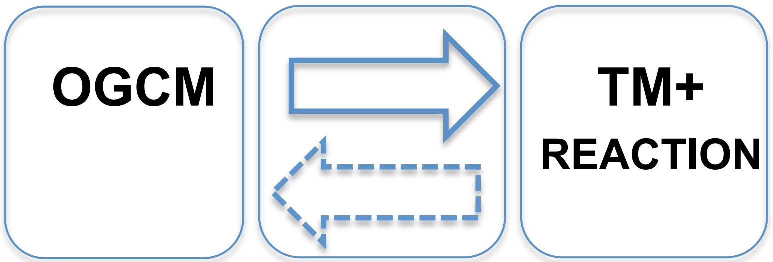
The best approach depends on the specific problem

Joeckel, 2012
Kerweg & Joeckel, GMD,2012

Linking OGCM → transport reaction equations

OFFLINE approach

ONLINE approach



Applications

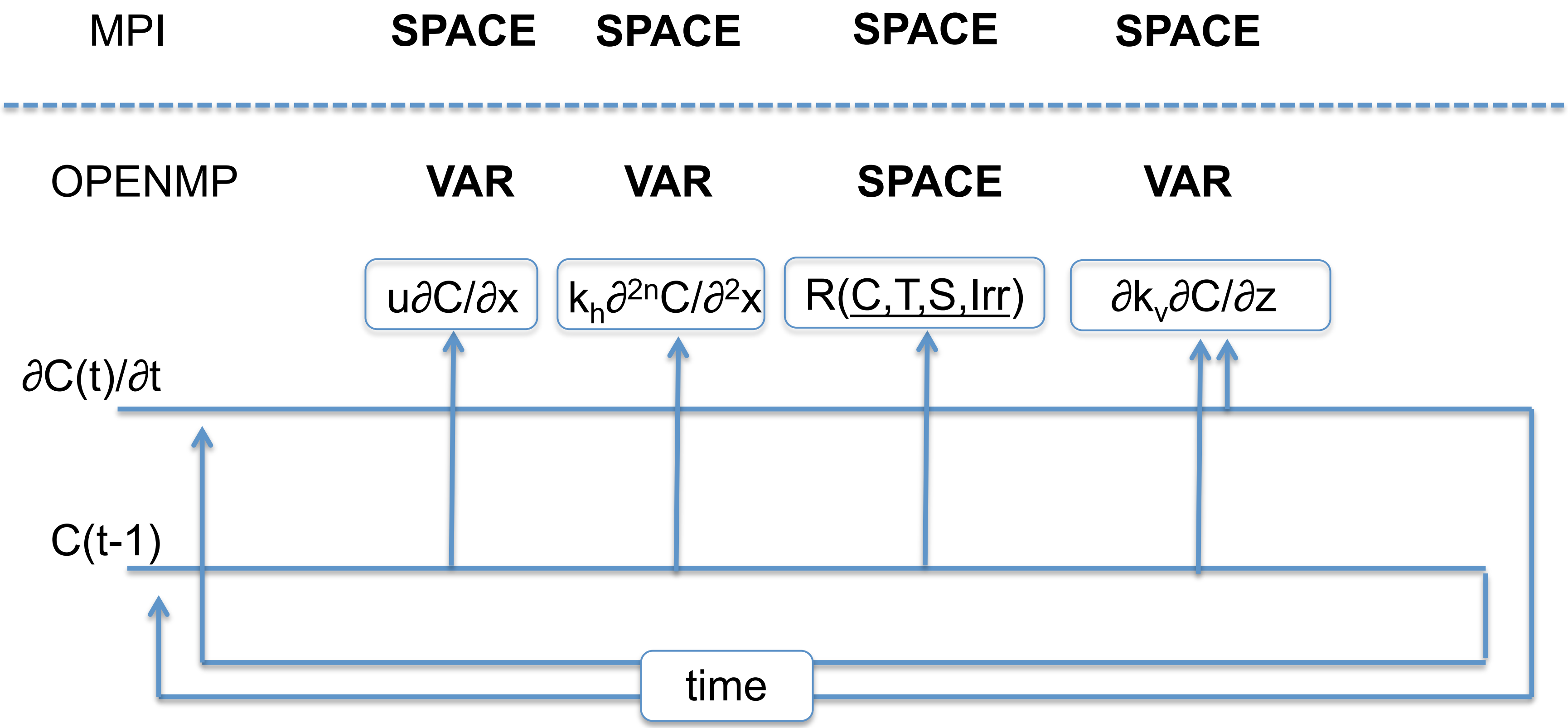
- ✓ intermediate processing
- ✓ Operational chain
- ✓ Climate simulations
- ✓ Sensitivity analisys

- ✓ MITgcm
- ✓ Nemo

… The OFFLINE APPROACH ...

OFFLINE - The transport component

Different approaches for parallellisation (SPACE → domain decomposition, VAR → parallelisation on biogeochemical variables)



Usually > 24 cores for a real applications

✓ To carry out those computations a lot of computational resources are required

✓Attention to optimization in order to carry out experiments in an acceptable time, and reducing I/O problems

✓ use of HPC facilities

Linking OGCM → transport reaction equations

OFFLINE approach

OGCM

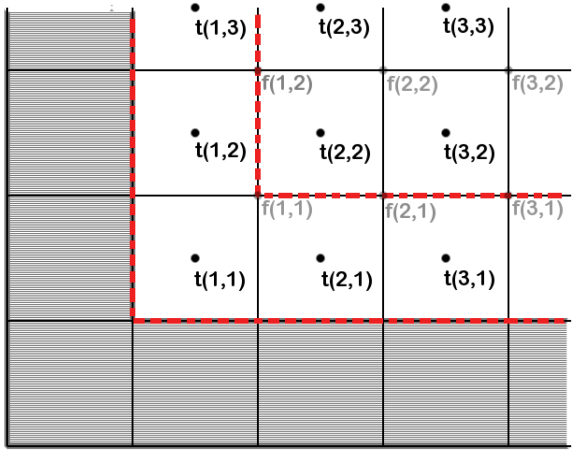
TM+
REACTION

Applications

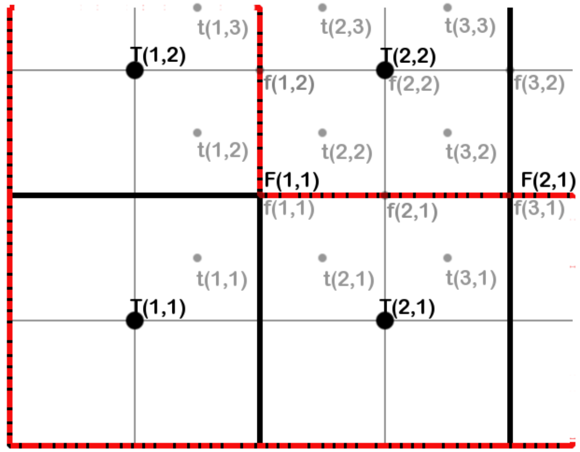
- ✓ **intermediate processing**
- ✓ Operational chain
- ✓ Climate simulations
- ✓ Sensitivity analisys

INTER-proc: Cell merge approach

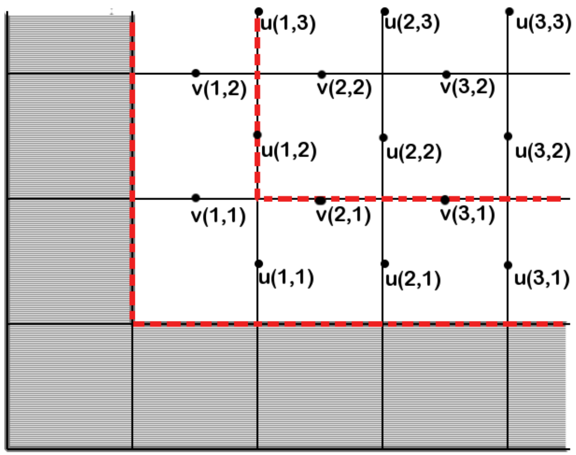
1/16° scalar fields



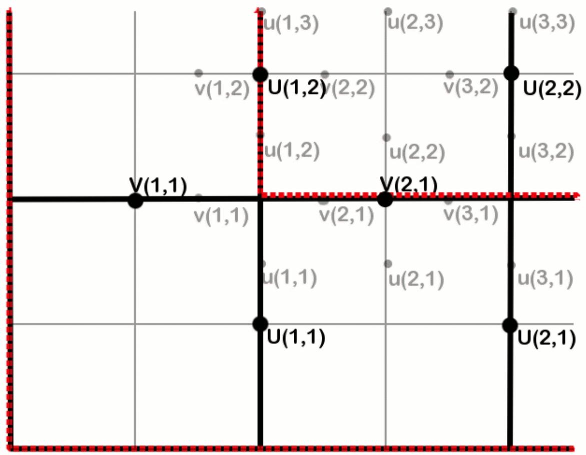
1/8° scalar fields



1/16° vectorial fields



1/8° vectorial fields

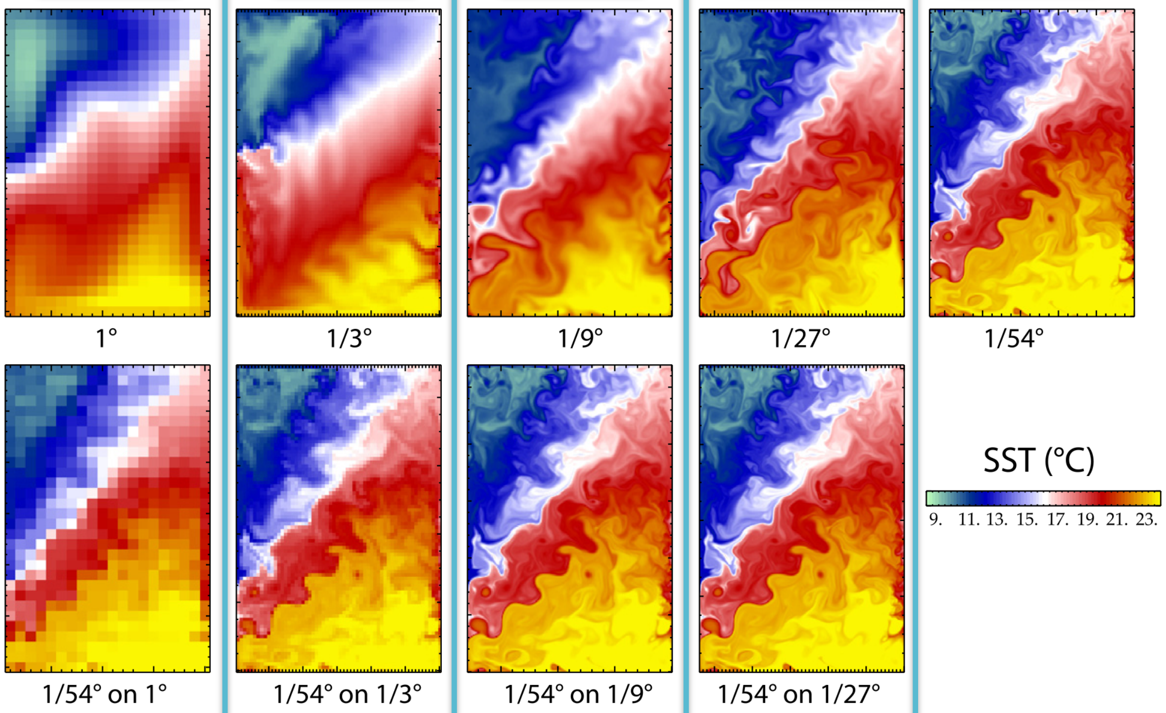


(Aumont et al., 1998; Lazzari et al.,2010; Levy et al., 2012)

Reduction in computational load

Resolution of the trasport OFFLINE model for temperature

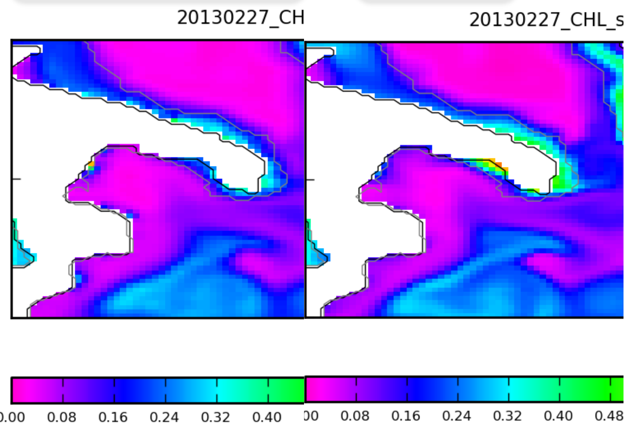
1°	1/3°	1/9°	1/27°	1/54°	← TM resolution



Levy et al., 2012

Physics requires higher resolution than tracers
Reduction in computational load, memory load, and storage space occupied

1/16° on 1/8°	1/16°



Coastal areas could be affected

Linking OGCM → transport reaction equations

OFFLINE approach

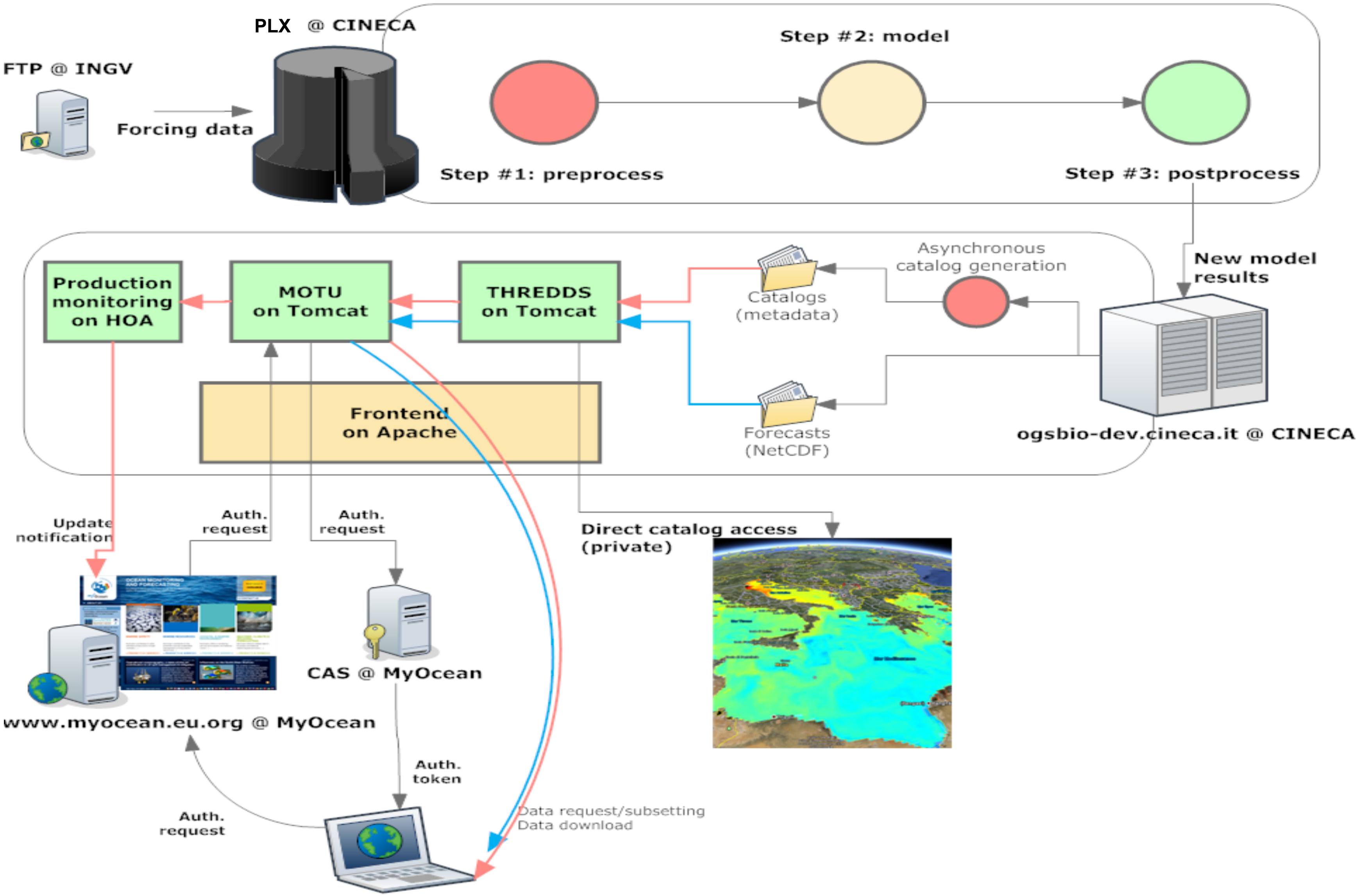
OGCM

TM+
REACTION

Applications

- ✓ intermediate processing
- ✓ **Operational chain**
- ✓ Climate simulations
- ✓ Sensitivity analisys

Operational Chain in OFFLINE mode

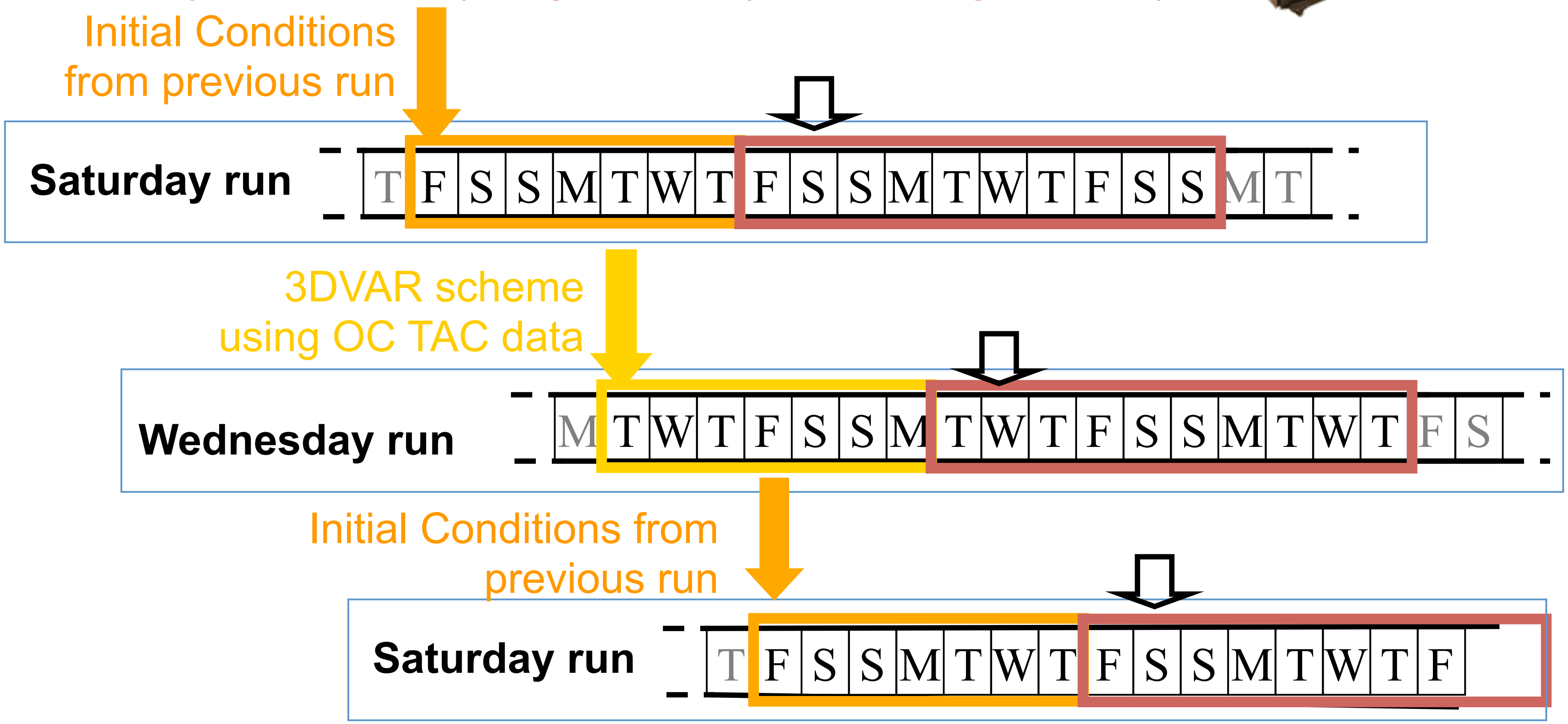


The operational chain

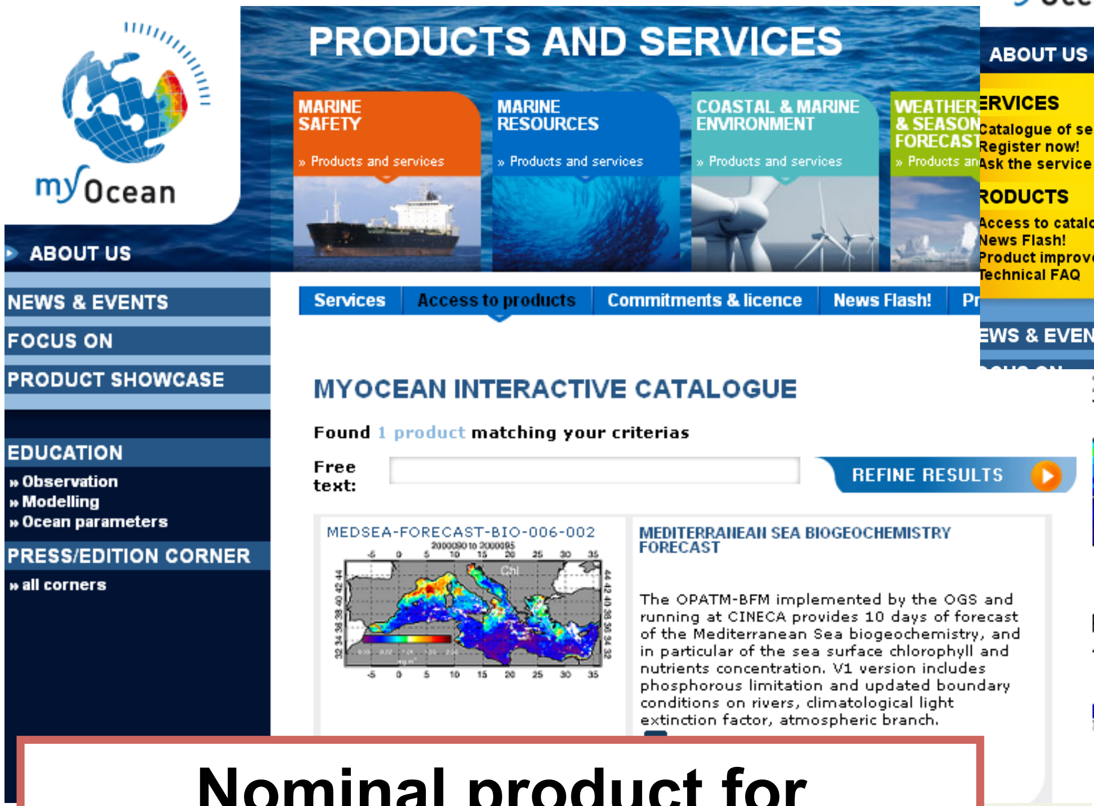
implementation in MyOcean MedFC

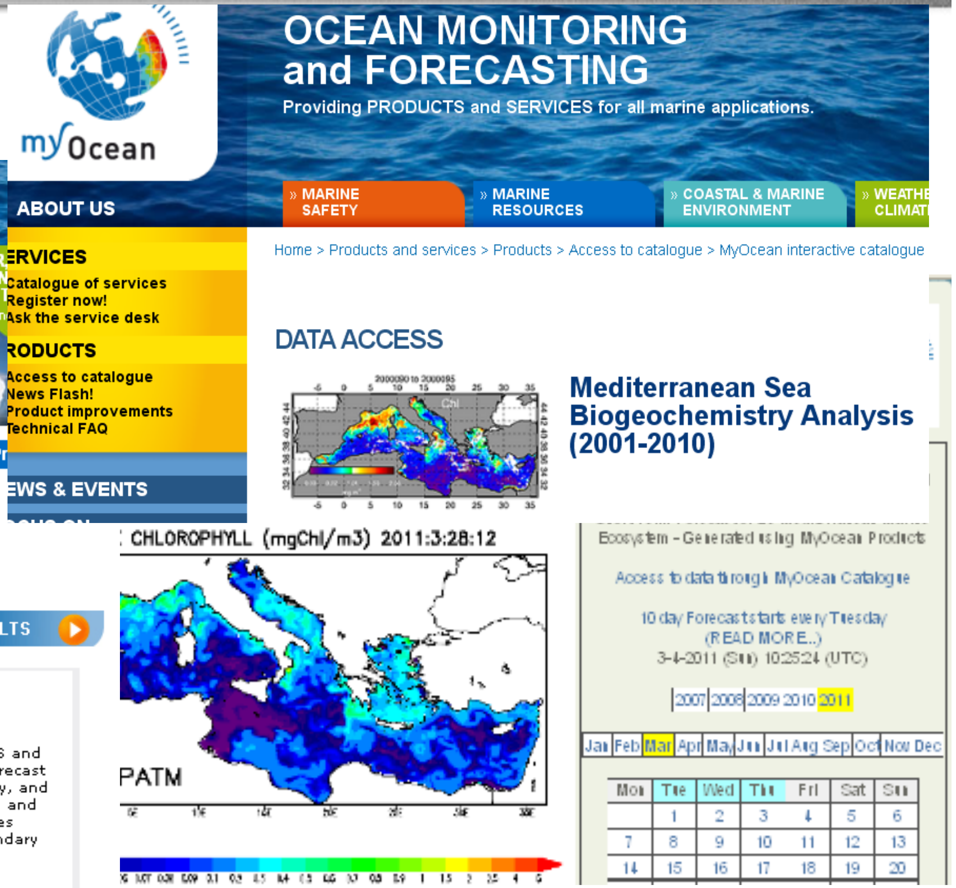
7 days of hindcast/analysis (using INGV physical forcing analysis and ICs via DA based on GOS-ISAC-CNR satellite chlorophyll)

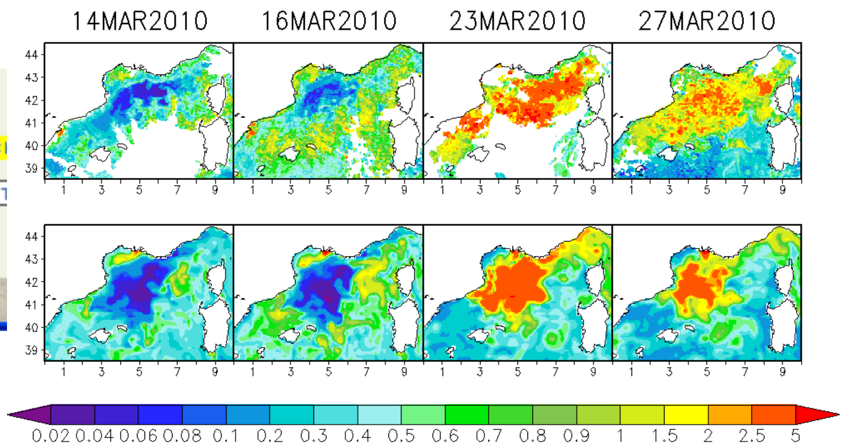
10 days of forecast (using INGV physical forcing forecast)



OGS products in MyOcean







Nominal product for biogeochemistry forecast in Med Sea + OGS web page + case studies + CalVal

http://data.ncof.co.uk/calval/

Linking OGCM → transport reaction equations

OFFLINE approach

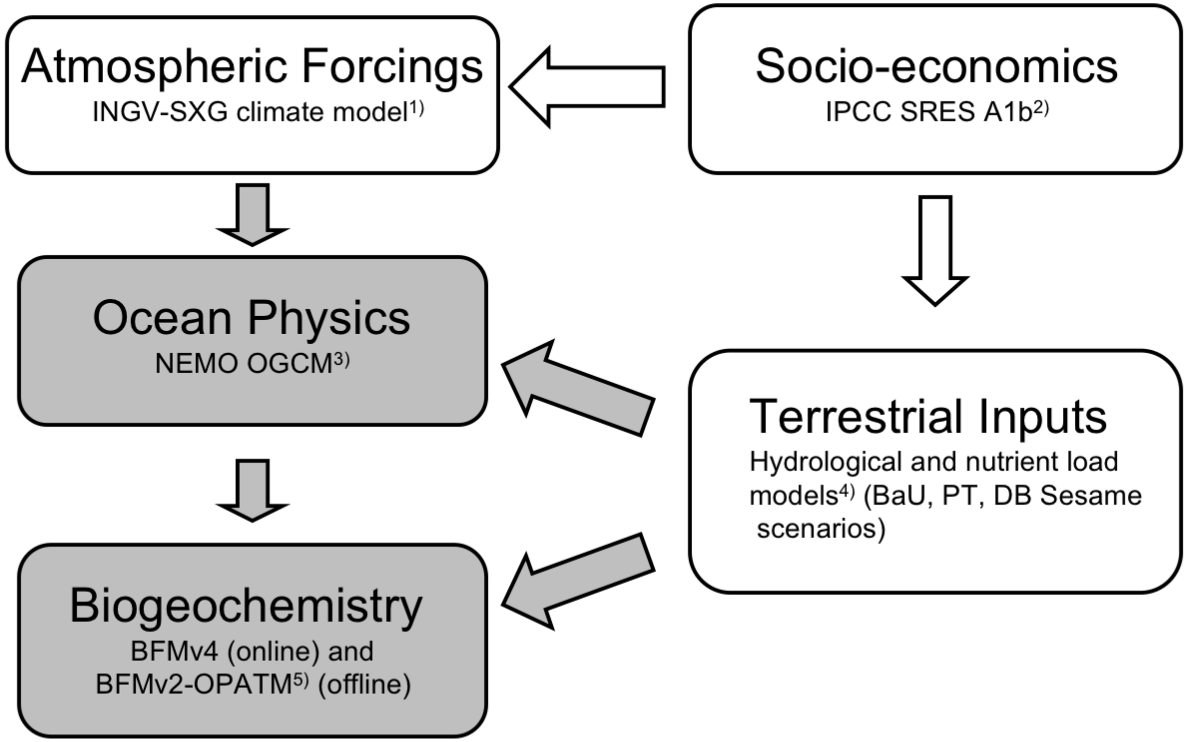
OGCM

TM+
REACTION

Applications

- ✓ intermediate processing
- ✓ Operational chain
- ✓ **Climate simulations**
- ✓ Sensitivity analisys

Simulations on climatic time scales



1) Gualdi et al. (2008); 2) Nakicenovic and Swart (2000); 3) Oddo et al (2009); 4) Ludwig et al. (2010); 5) Lazzari et al. (2014)

- Increase of carbon rates both production (GPP) and community respiration (RSP)
- Increase of dissolved semi-labile carbon
- Reduction in biomass

MEDITERRANEAN BASIN

	20C	A1B-BaU	A1B-PT	A1B-DB
GPP	0.66	0.044	0.047	0.029
RSP	0.65	0.044	0.048	0.030
NPP	0.36	0.032	0.036	0.015
NCP	0.01	-0.001	-0.033	-0.064
DSL	0.96	0.038	0.044	0.035
BIO	4.12	-0.046	-0.043	-0.056

WESTERN BASIN

	20C	A1B-BaU	A1B-PT	A1B-DB
GPP	0.81	0.023	0.017	0.011
RSP	0.80	0.023	0.016	0.013
NPP	0.46	0.009	0.001	-0.006
NCP	0.01	0.050	0.019	-0.002
DSL	1.02	0.030	0.027	0.025
BIO	4.97	-0.070	-0.074	-0.076

EASTERN BASIN

	20C	A1B-BaU	A1B-PT	A1B-DB
GPP	0.58	0.061	0.073	0.044
RSP	0.56	0.063	0.076	0.046
NPP	0.30	0.053	0.067	0.034
NCP	0.01	-0.035	-0.068	-0.104
DSL	0.93	0.045	0.056	0.042
BIO	3.63	-0.027	-0.018	-0.039

Linking OGCM → transport reaction equations

OFFLINE approach

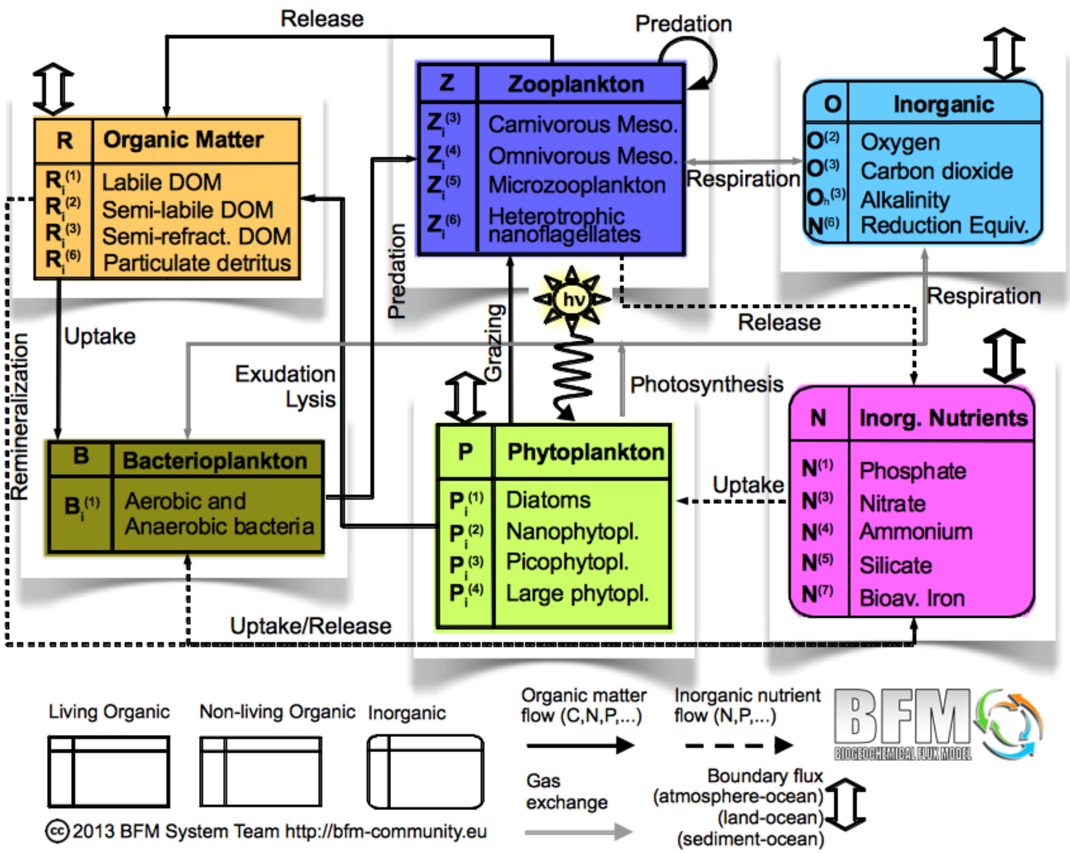
OGCM

TM+
REACTION

Applications

- ✓ intermediate processing
- ✓ Operational chain
- ✓ Climate simulations
- ✓ **Sensitivity analisys**

Sensitivity to parameters uncertainty



9 PFT with **~ 20** physiological processes each

~200 parameters control the kinetic of the processes

What is the uncertainty in the model results?

Morris's method

a base vector is randomly chosen within the parameter space

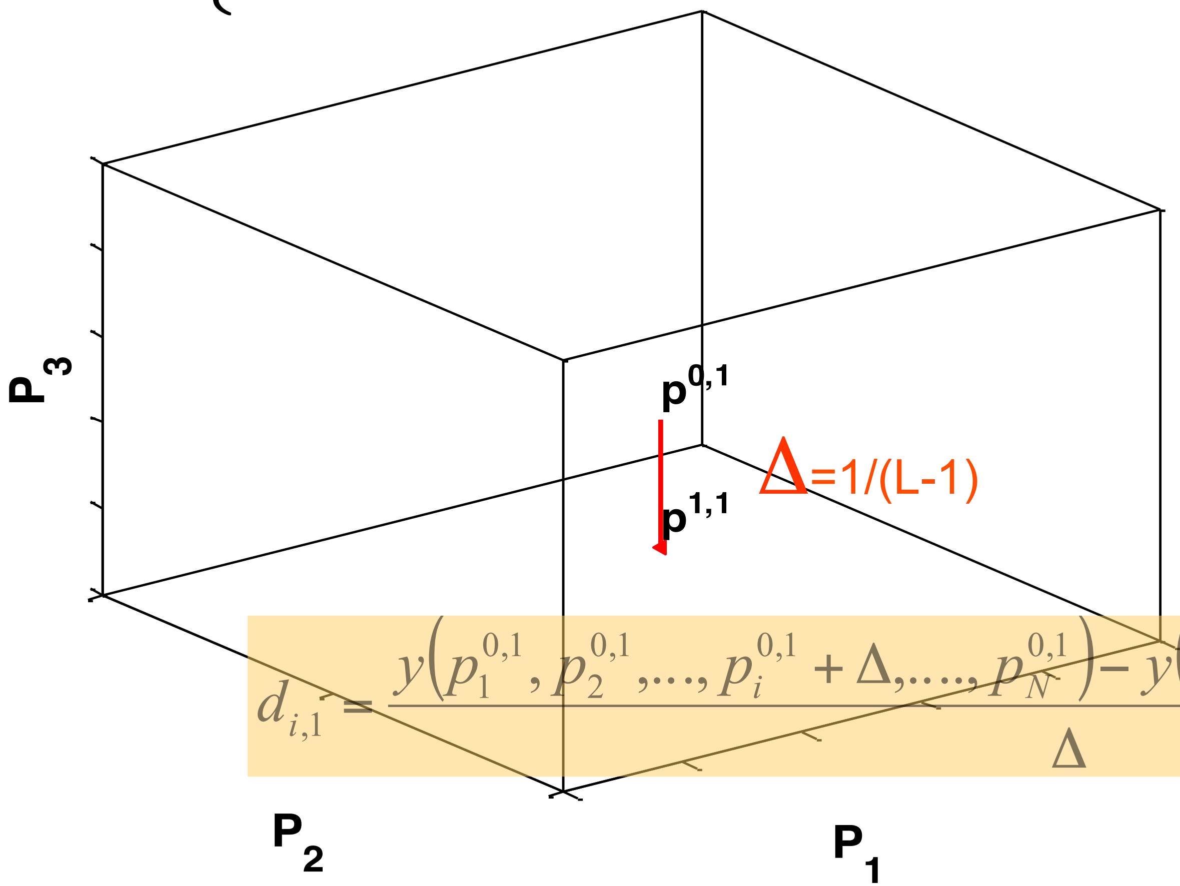
$$\mathbf{P^{0,1}} = \left\{ P_1^{0,1}, P_2^{0,1}, \ldots, P_i^{0,1}, \ldots, P_N^{0,1} \right\}$$

one parameter is randomly chosen and varied of Δ

$$\mathbf{P^{1,1}} = \left\{ P_1^{0,1}, P_2^{0,1}, \ldots, P_i^{0,1} + \Delta, \ldots, P_N^{0,1} \right\}$$

elementary effects on model output

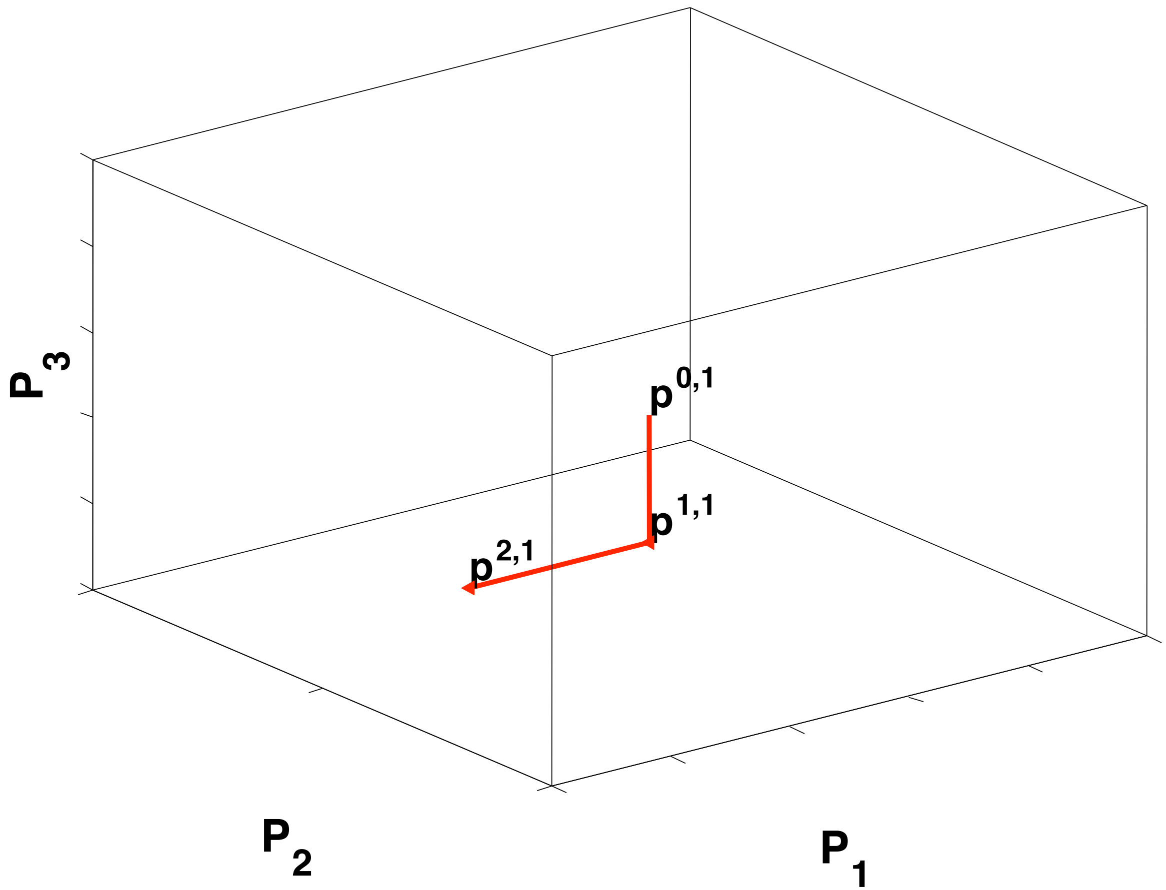
$d_{1,3}$



$$d_{i,1} = \frac{y\left(p_1^{0,1}, p_2^{0,1}, \ldots, p_i^{0,1} + \Delta, \ldots, p_N^{0,1}\right) - y\left(p_1^{0,1}, p_2^{0,1}, \ldots, p_i^{0,1}, \ldots, p_N^{0,1}\right)}{\Delta}$$

How it works … 1

a second parameter is randomly chosen to be varied of Δ



Elementary effects on model output

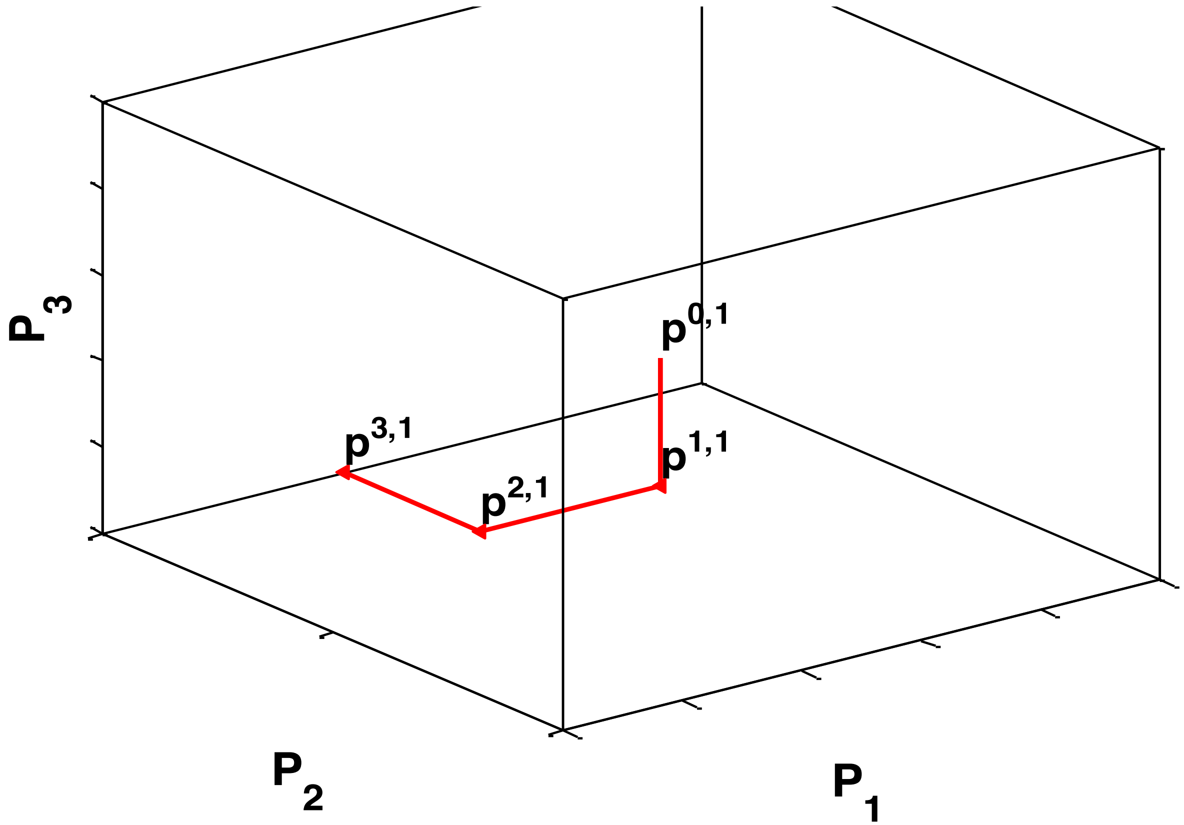
$d_{1,1}$ $d_{1,3}$

How it works …2

the step is repeated until all of the parameters have been varied
a trajectory on the space of the parameter is then performed

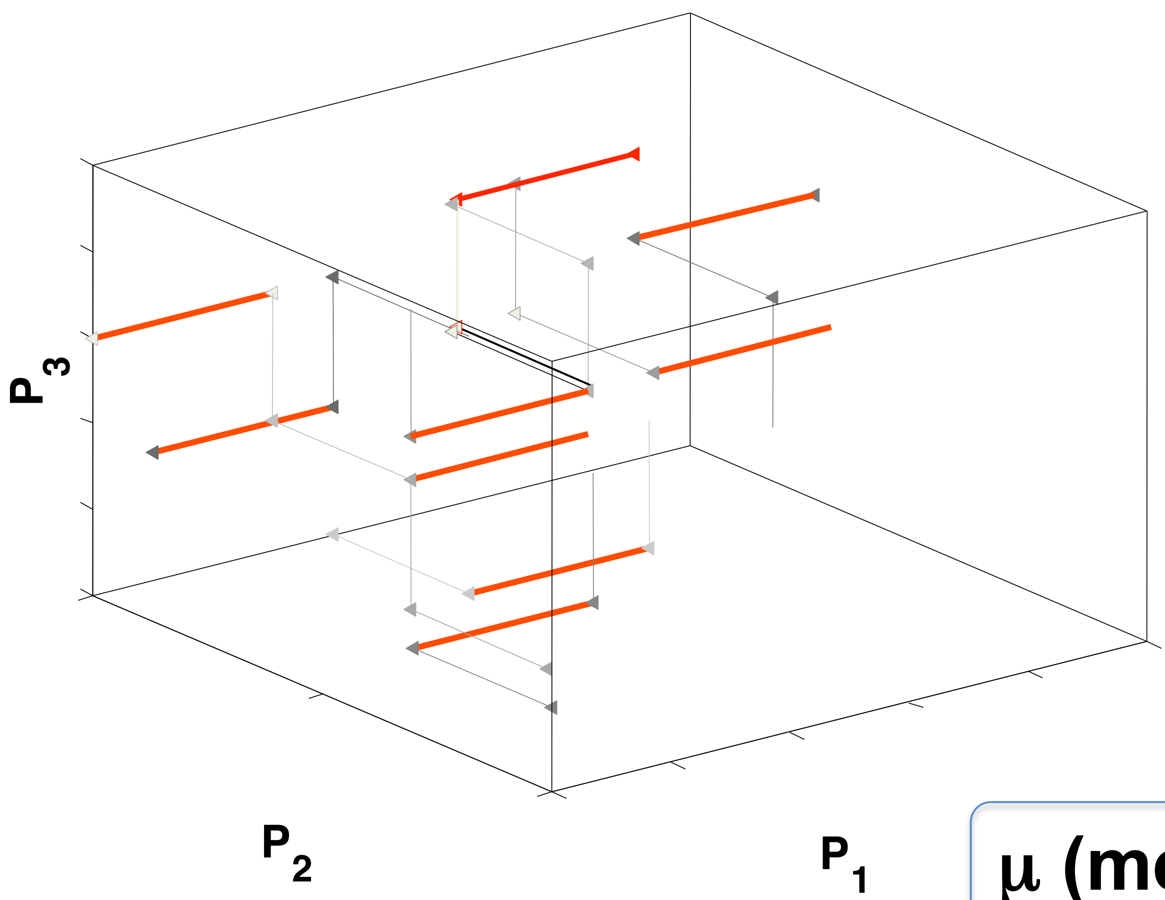
one elementary effect for each param.

$d_{1,1}$ $d_{1,2}$ $d_{1,3}$



How it works …3

distribution of elem. effects on the model output due to variations of the first parameter



elementary effects on model output

$d_{1,1}$	$d_{1,2}$	$d_{1,3}$
$d_{2,1}$	$d_{2,2}$	$d_{2,3}$
$d_{3,1}$	$d_{3,2}$	$d_{3,3}$
$d_{4,1}$	$d_{4,2}$	$d_{4,3}$
$d_{5,1}$	$d_{5,2}$	$d_{5,3}$
$d_{6,1}$	$d_{6,2}$	$d_{6,3}$
$d_{7,1}$	$d_{7,2}$	$d_{7,3}$
$d_{9,1}$	$d_{9,2}$	$d_{9,3}$
$d_{9,1}$	$d_{9,2}$	$d_{9,3}$
$d_{10,1}$	$d_{10,2}$	$d_{10,3}$

μ (mean), σ (st.dev.)

To carry out this experiment …

… on a 3D implementation of BFM

- ✓ sensititity of 165 parameters on 122 variables in 51 different conditions (trajectories)
- ✓ this results in about 8500 simulations
- ✓ 21 Mh cpu
- ✓ 24 Milions of files produced
- ✓ 54 TB of space
- ✓ using vectorization of 3D files –without that we would have 240 TB !

Some considerations …

Model I/O

The run at BlueGeneQ facility of CINECA, where I/O is a bottleneck. Vectorization has been a crucial solution to:

- ✓ save disk space
- ✓ Save time both in simulation and in postproc
- ✓ Avoid zip/unzip (time consuming)

POSTPROC

Output must be reduced size in order

- ✓ to be stored and moved elsewhere
- ✓ to investigate easily on local servers the relationships parameter-variable

1 file for simulation (~10000 files)

1TB of space → that's enough for science

GSENSMED - doing the job

Simulation

Parallelism Level 1 : Pure MPI Code, Mediterranean Sea divided in 128 tasks

Parallelism Level 2 : Since the BlueGeneQ minimal requirement is 2048 cores, we used the SUBBLOCK task-farm technique. A job launch 16 different mpirun using a subset of cores

Postproc

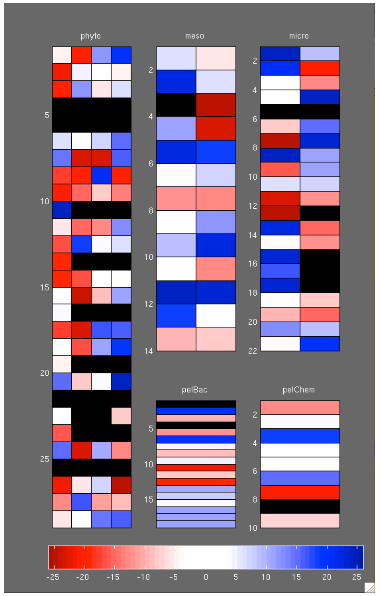
With job of 2048 cores we process results of 165 simulations x 122 variables, i.e. we do 20000 times the same statistical work with the task-farm technique.

Solution adopted to remain on filesystem : 6D array

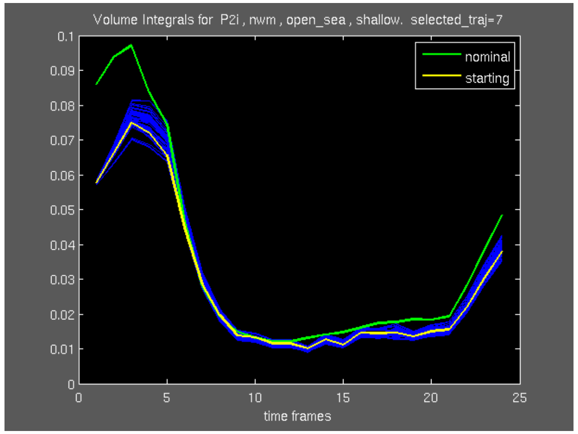
PROFILES [var, time, subbasin, distance from coast, z, kind of statistic]

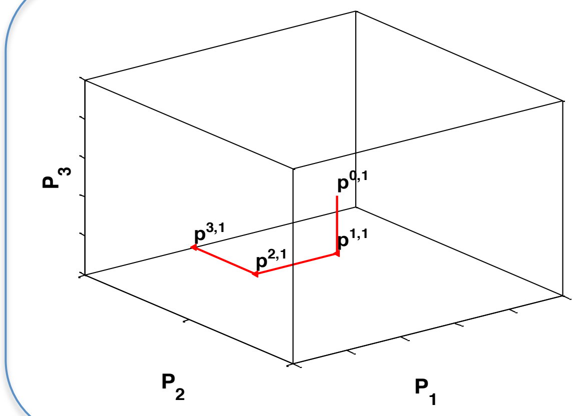
No communication between ranks, every rank writes its file, then a serial collector writes the final file

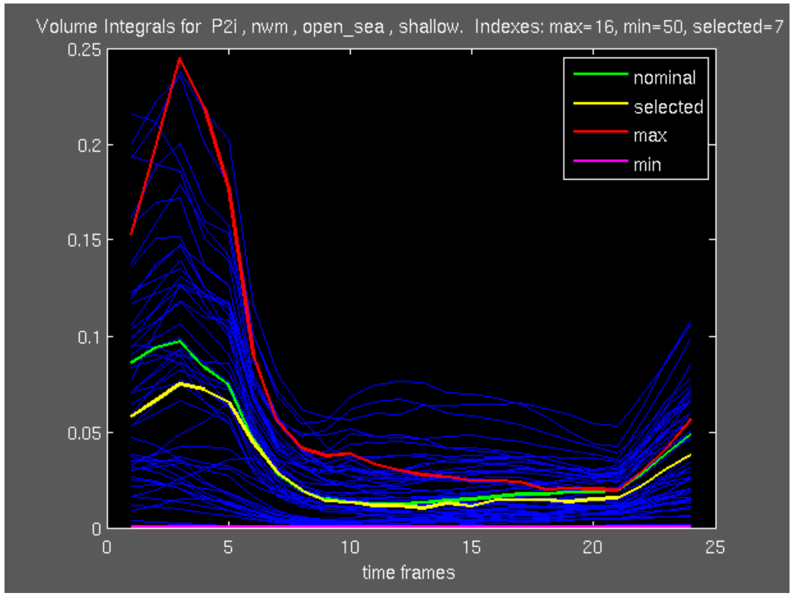
GSENSMED - Results

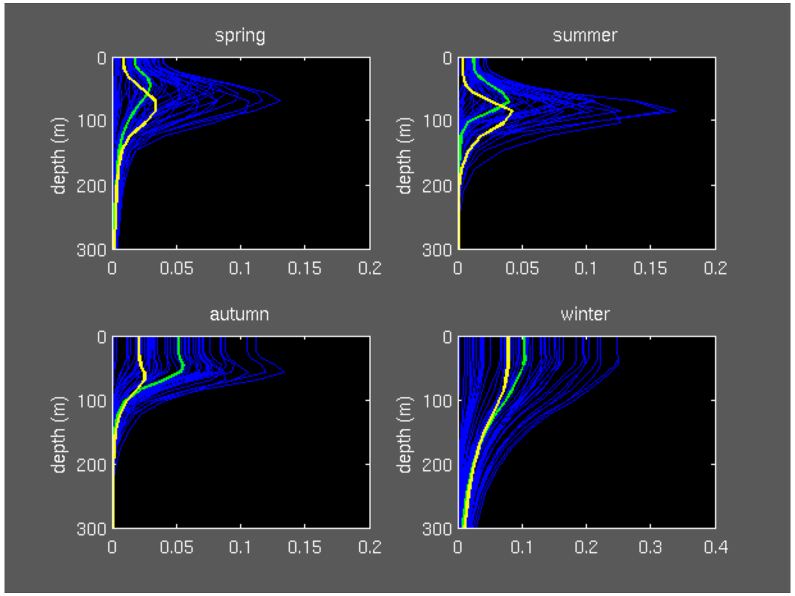


← Perturbation map of the namelist file



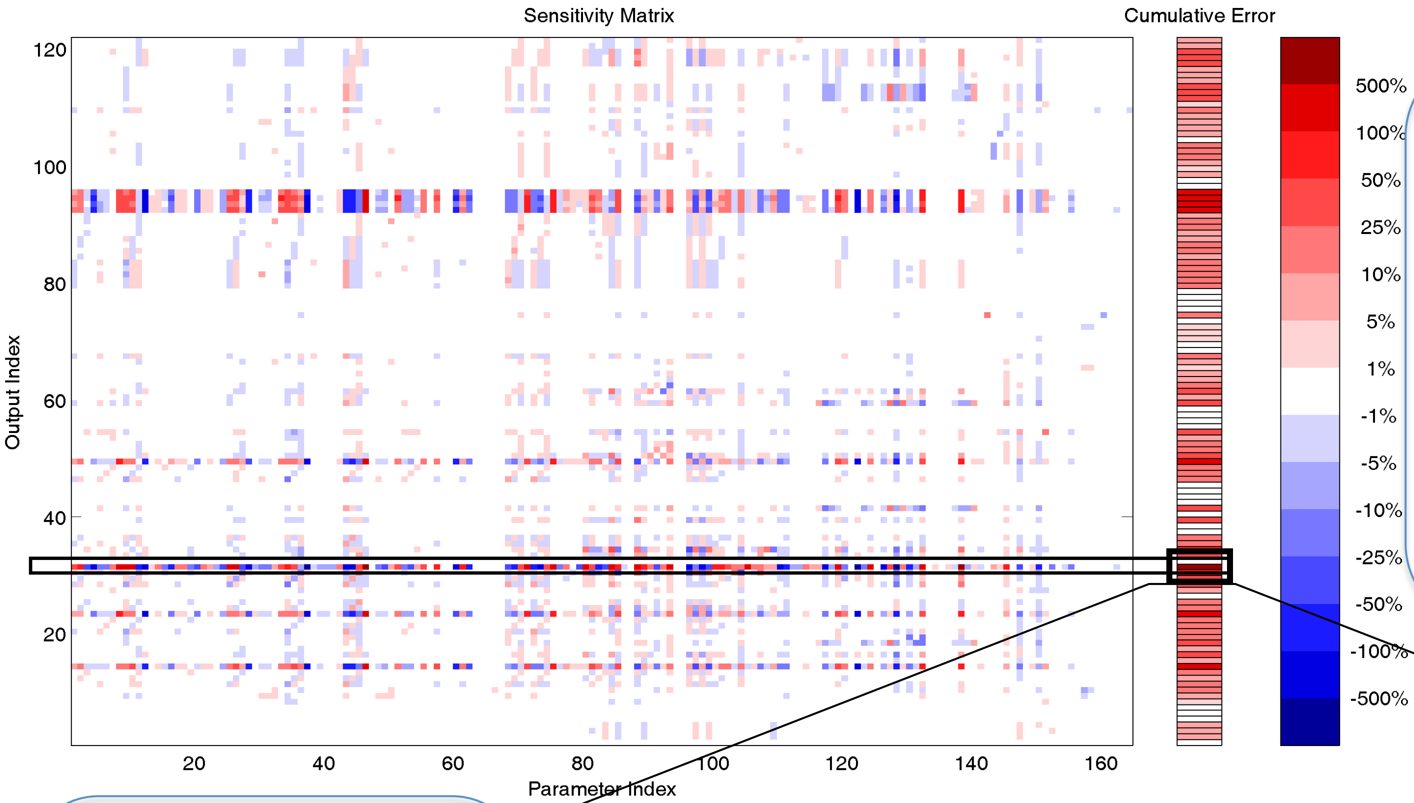






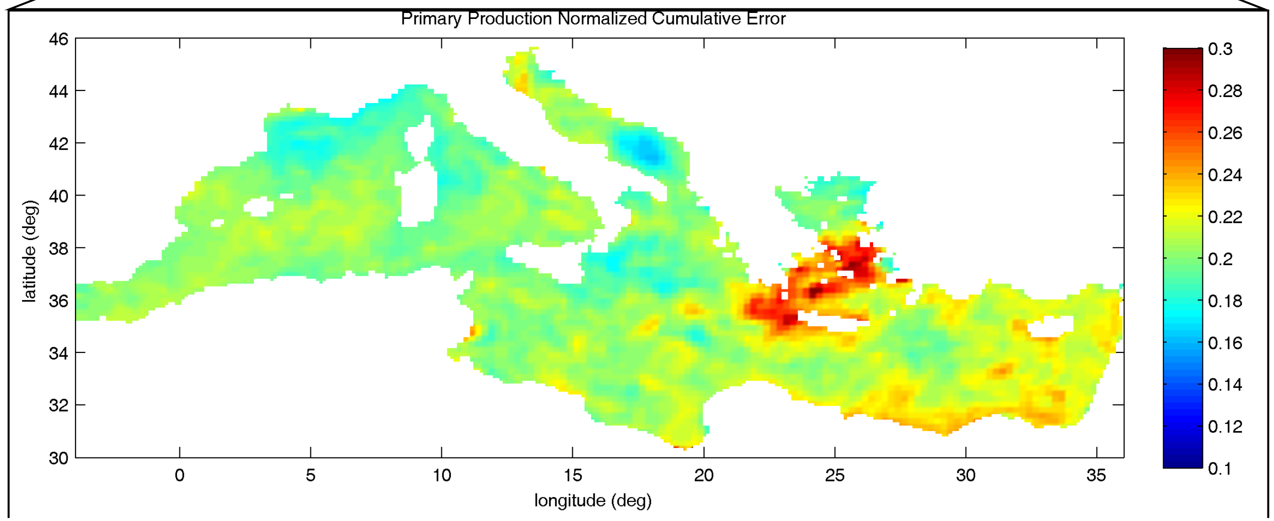
Results for the initial trajectories (51)

GSENSMED: Global sensitivity experiment



Sensitivity of each model output (prognostic and diagnostic variables) as function of a 5% uncertainty on each model parameter
Global model uncertanty

Map of annual average of the normalized cumulative error associated to primary production



A 5% uncertainty in the value of each of the hundreds model parameters, with an estimate of the cumulative effect of such uncertainty.

Globally, the uncertainty is higher (up to 300%) in measures related to higher trophic level organisms, such as carnivorous mesozoopkanton, and lower for those related to planktonic primary producers

Several outputs present pretty large uncertainty, but –globally- the cumulative effect of potential uncertainty in hundreds of model parameters add up to an error which is smaller than 25% for the majority of the parameters.

In detail, the uncertainty associated to politically very relevant quantities, such as total primary productivity (a first proxy of sea fertility) or air-sea CO_2 fluxes (a first proxy of carbon sequestration) are about 5%.

The OFFLINE coupling approach

Advantages

✓Reduction of the computational load especially needed in application with O(100) tracers to be advected (cell merge approach);

✓ capability to link with different OGCM (Operational or climate);

✓ the Offline approach simplify the task to carry out sensitivity analysis of complex 3D coupled models;

✓Results are essentially equivalent to TM forced with higher resolution model (Levy et al., 2013).

Disadvantages

✓ formulation of the Transport Model must be sufficiently congruent with the formulation of the OGCM;

✓ Loose high frequency effects

✓ feedbacks of biology versus physics are not represented;

✓disk space needed to store forcings (for 140 yrs~3.2Tb).

… COUPLING THE REACTOR IN A MODULAR WAY…

Shading effect of bio on radiation

A decrease in chl from 1.0 to 0.1 mg m^{-3} results in a 10 fold increase of net solar flux at 20 m

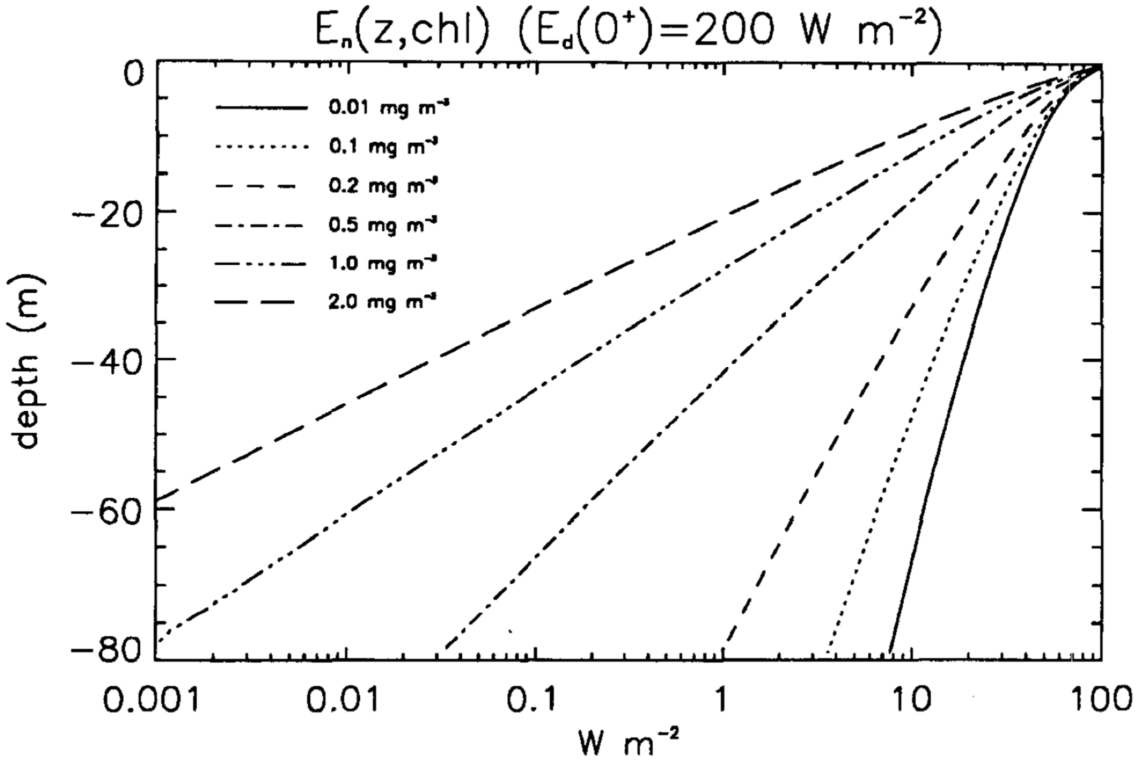


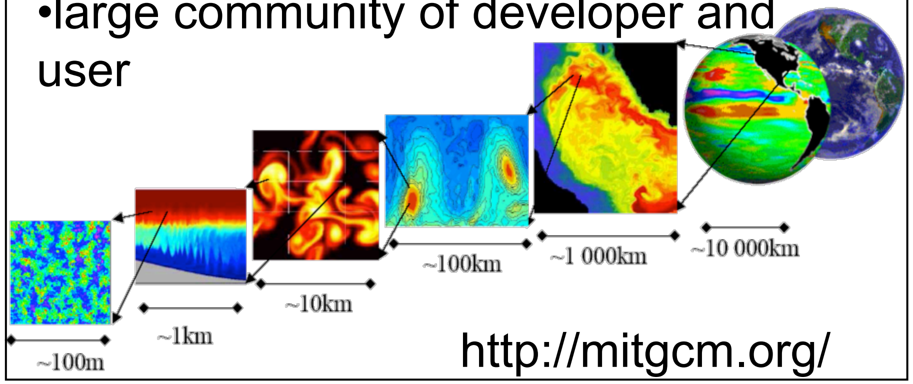
FIG. 3. Vertical profiles of net irradiance for an incident solar flux of 200 W m^{-2} and chlorophyll *a* concentrations ranging from 0.01 to 2.0 mg m^{-3} after spectral integration from 300 to 750 nm.

Ohlmann et al, 1996

COUPLED PHYSICAL-BIOGEOCH. MODEL

MITgcm: M.I.T. General Circulation Model

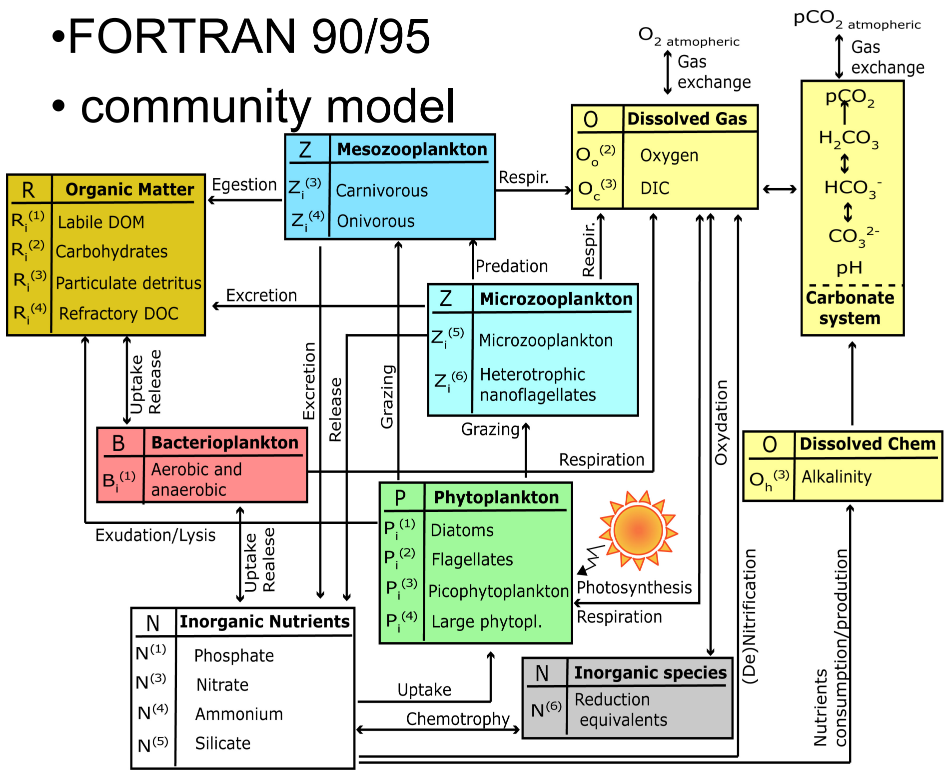
- designed to study both atmospheric and oceanic phenomena, at various scales (from meter to global)
- includes non-hydrostatic capability
- adopts the KPP vertical turbulence parametrization
- adopts a finite volume technique
- developed to perform efficiently on a wide variety of computational platforms
- FORTRAN 77
- large community of developer and user



http://mitgcm.org/

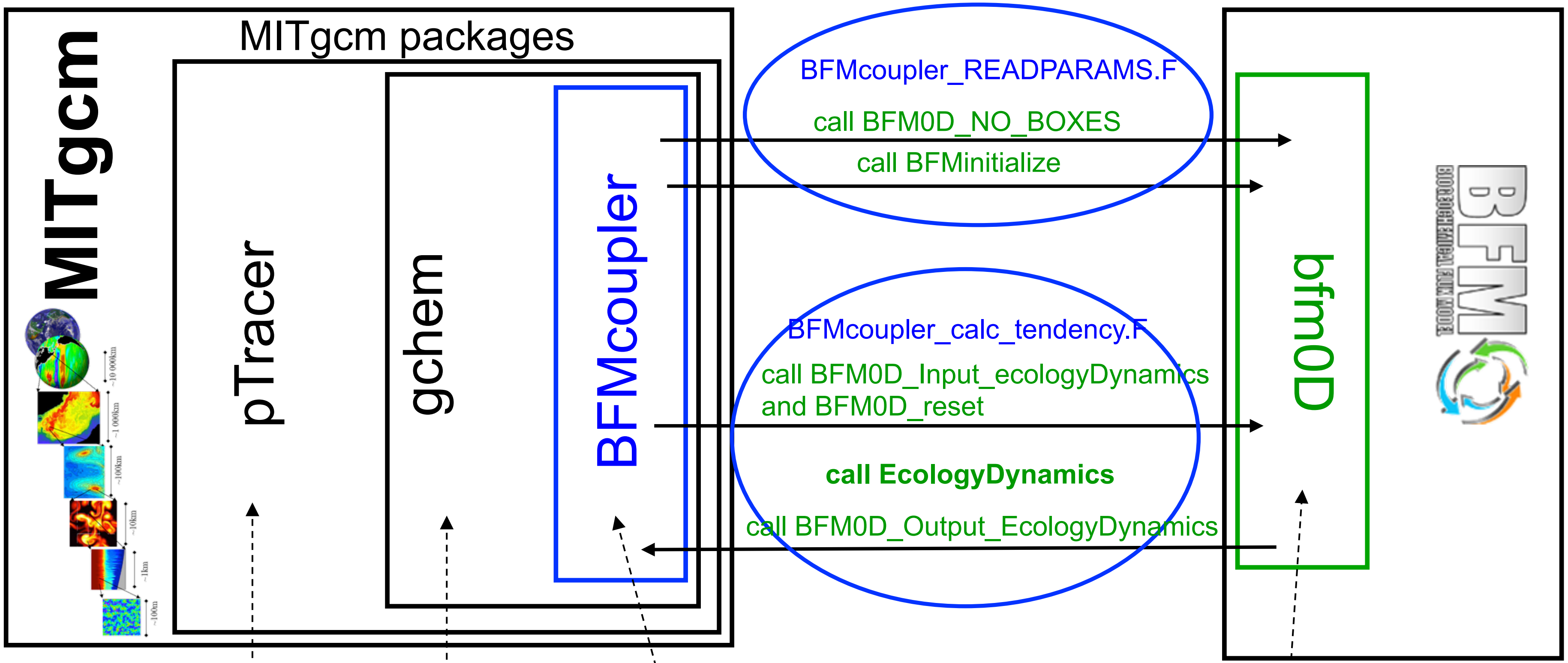
BFM Biogeochemical flux model

- Medium complexity food web
- carbonate system
- multi-nutrients (C, N, P, Si, Chla, Fe)
- modular code
- FORTRAN 90/95
- community model



http://bfm-community.eu/

Coupling MITgcm-BFM



BC and IC of tracers
Compute tendency for
transport (Adv&diff)

Generic coupler for tracers
call BFMcoupler routines
(read params, external
fields,initialization fields, and
diagnostics)

Allocation of memory on
0D configuration and
interface with main_BFM

Load and memory store of environmental local
variables (P and N Deposition, atmpCO2, atmP, atmWind,
surfPAR, PAR and light extinction factor) and
biogeochemical tendency of tracers

… Some technical problems ...

(to introduce the lab session)

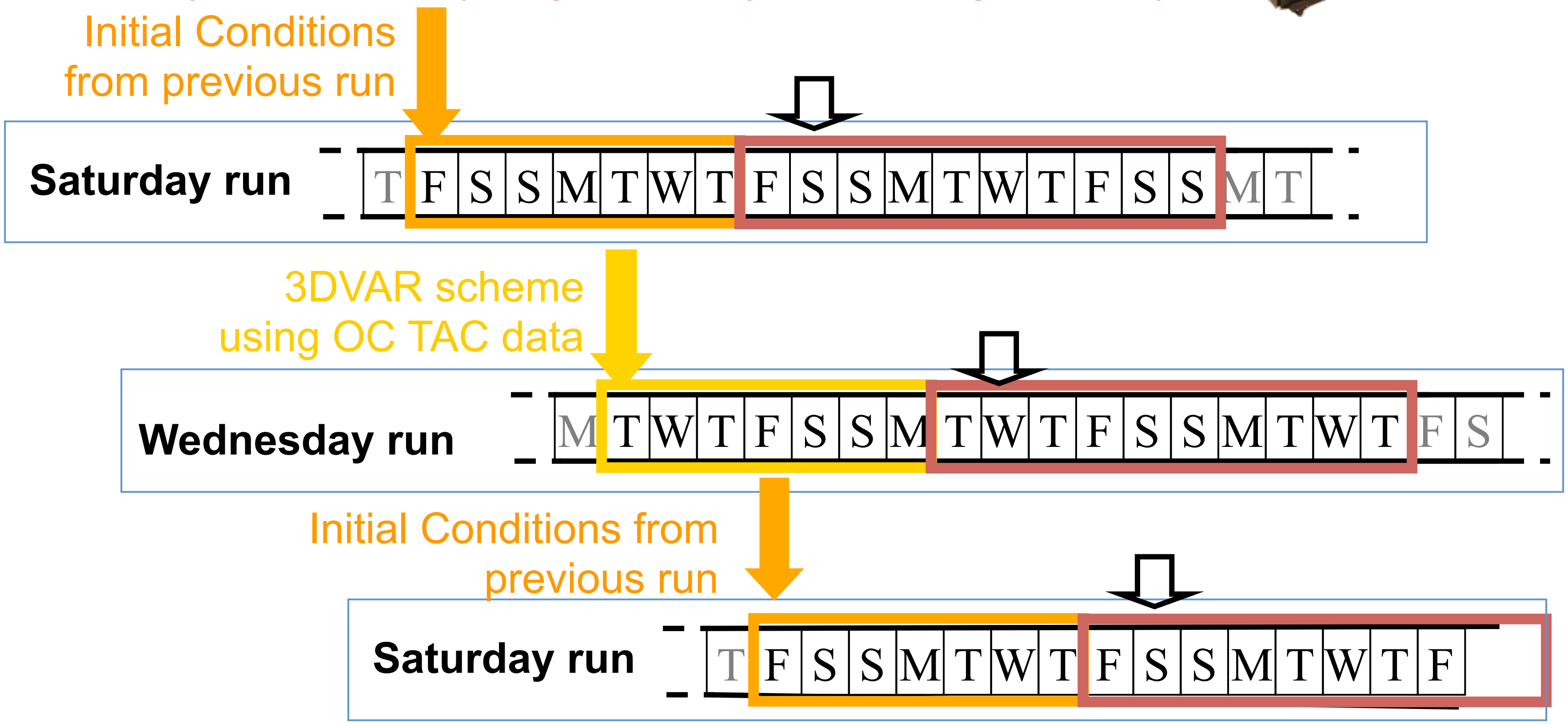
- ✓ Is it possibile to couple a particular complex postproc procedure in the main model algorithm?
- ✓ Is it possible to threat the I/O procedures as a separate model to optmize the model performance?
- ✓ … coupling a model with itself …

Including the DA in the model framework

implementation in MyOcean MedFC

7 days of hindcast/analysis (using INGV physical forcing analysis and ICs via DA based on GOS-ISAC-CNR satellite chlorophyll)

10 days of forecast (using INGV physical forcing forecast)



Data Assimilation scheme (DA)

3D variational approach

- Cost function to obtain the innovation

Dobricic and Pinardi, 2008

$$J(\delta \mathbf{x}_k) = \delta \mathbf{x}_k^{\ T} \mathbf{B}_k^{-1} \delta \mathbf{x}_k + (\mathbf{d}_k - \mathbf{H}_k \delta \mathbf{x}_k)^T \mathbf{R}_k^{-1} (\mathbf{d}_k - \mathbf{H}_k \delta \mathbf{x}_k)$$

- Model error covariance decomposition

$$\mathbf{B} = \mathbf{V}^T \mathbf{V} \qquad \delta \mathbf{x} = \mathbf{V} \mathbf{v}$$

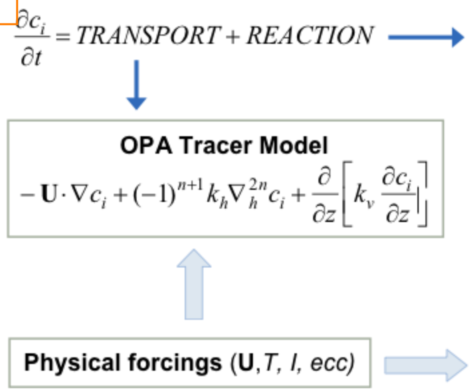
- Solution in the control space **v**

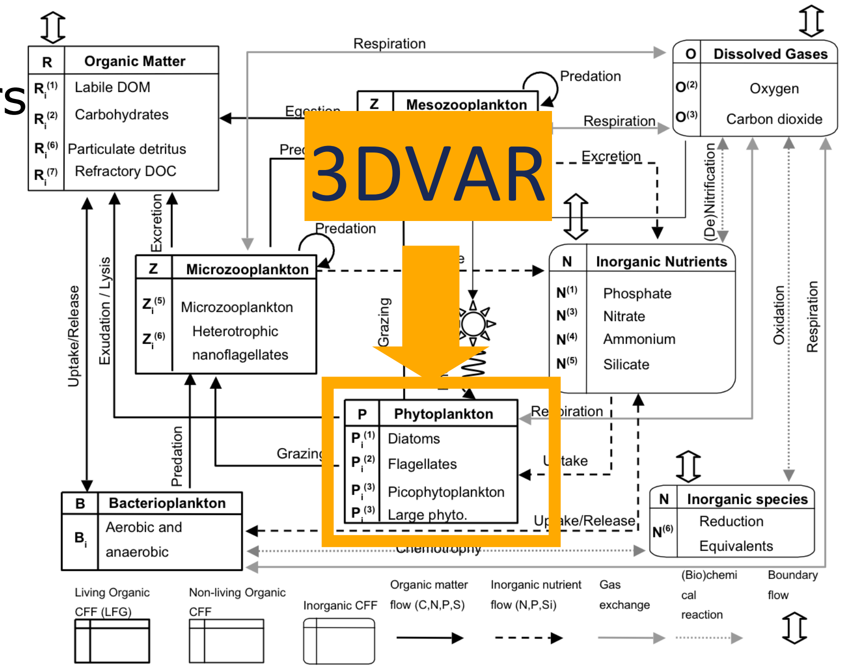
$$J(\mathbf{v}) = \mathbf{v}^{\mathrm{T}} \mathbf{v} + (\mathbf{d} - \mathbf{HVv})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{d} - \mathbf{HVv})$$

- **V** decomposed into a series of operators

$$\mathbf{V} = \mathbf{V}_{\mathbf{b}} \mathbf{V}_{\mathbf{h}} \mathbf{V}_{\mathbf{v}}$$

Teruzzi et al., 2014



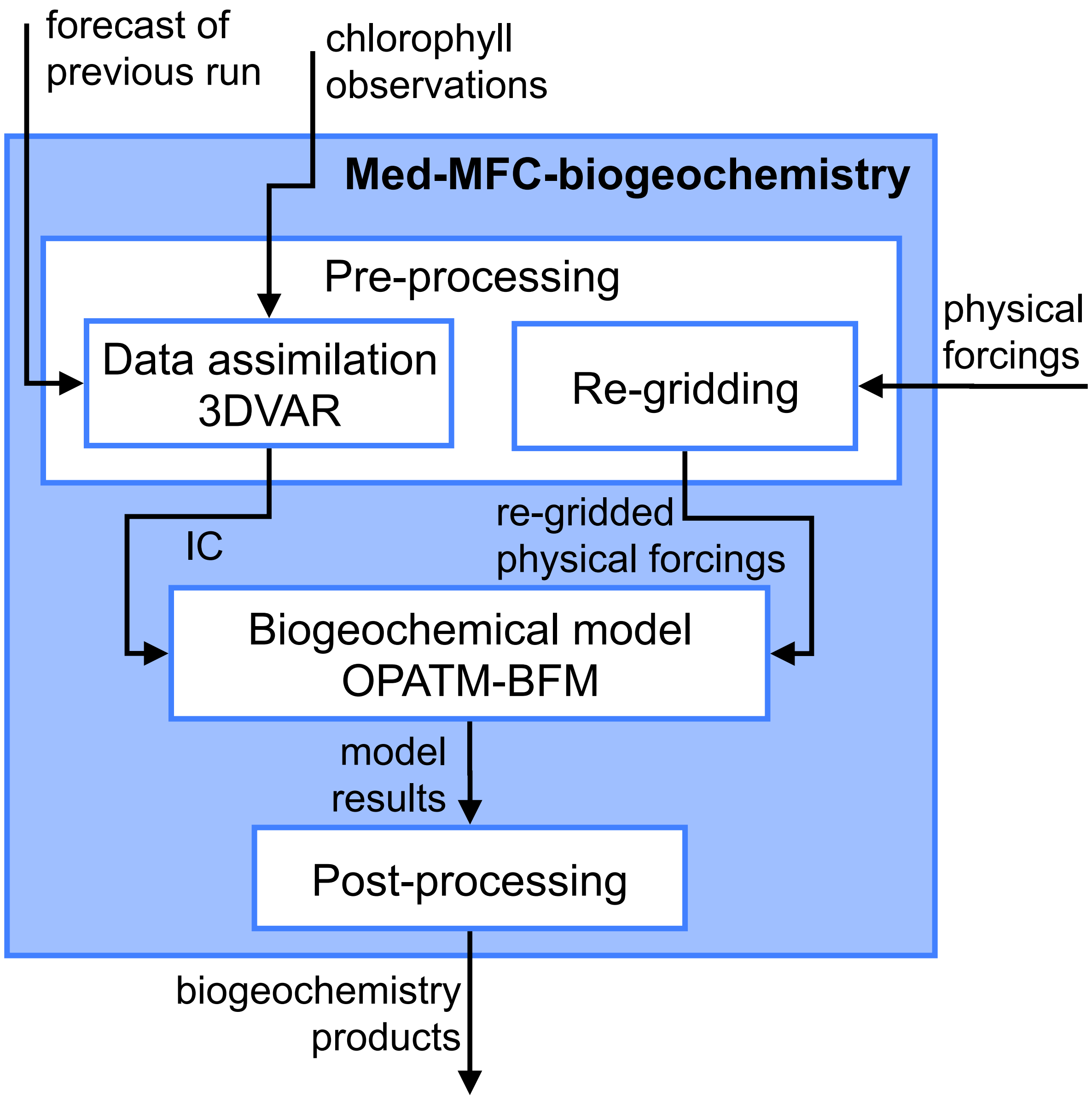


COUPLING THE DA WITH THE MODEL

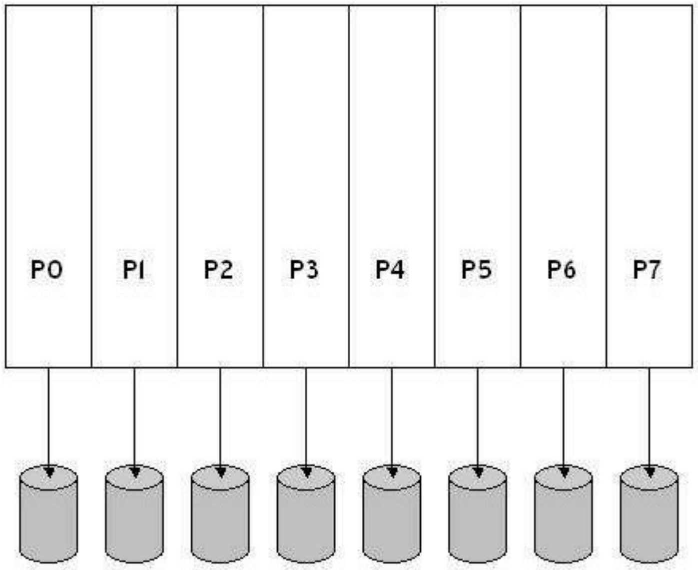
✓ To perfom a reanalys long sequences of run followed by assimilation

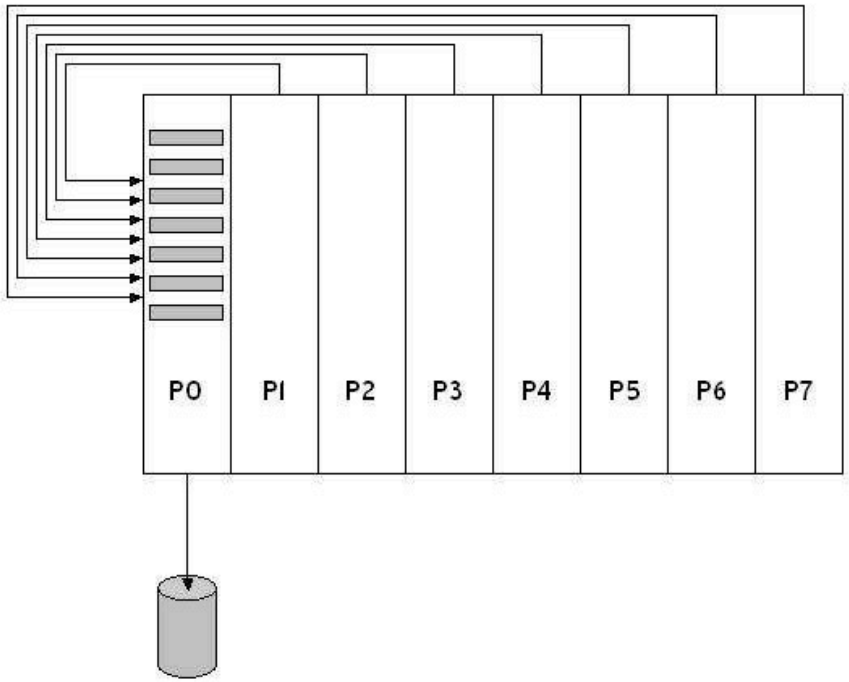
✓To avoid reinitialization of the model coupling of the transport and DA model

✓ If both are parallel can they "coexist in a single MPI framework"?



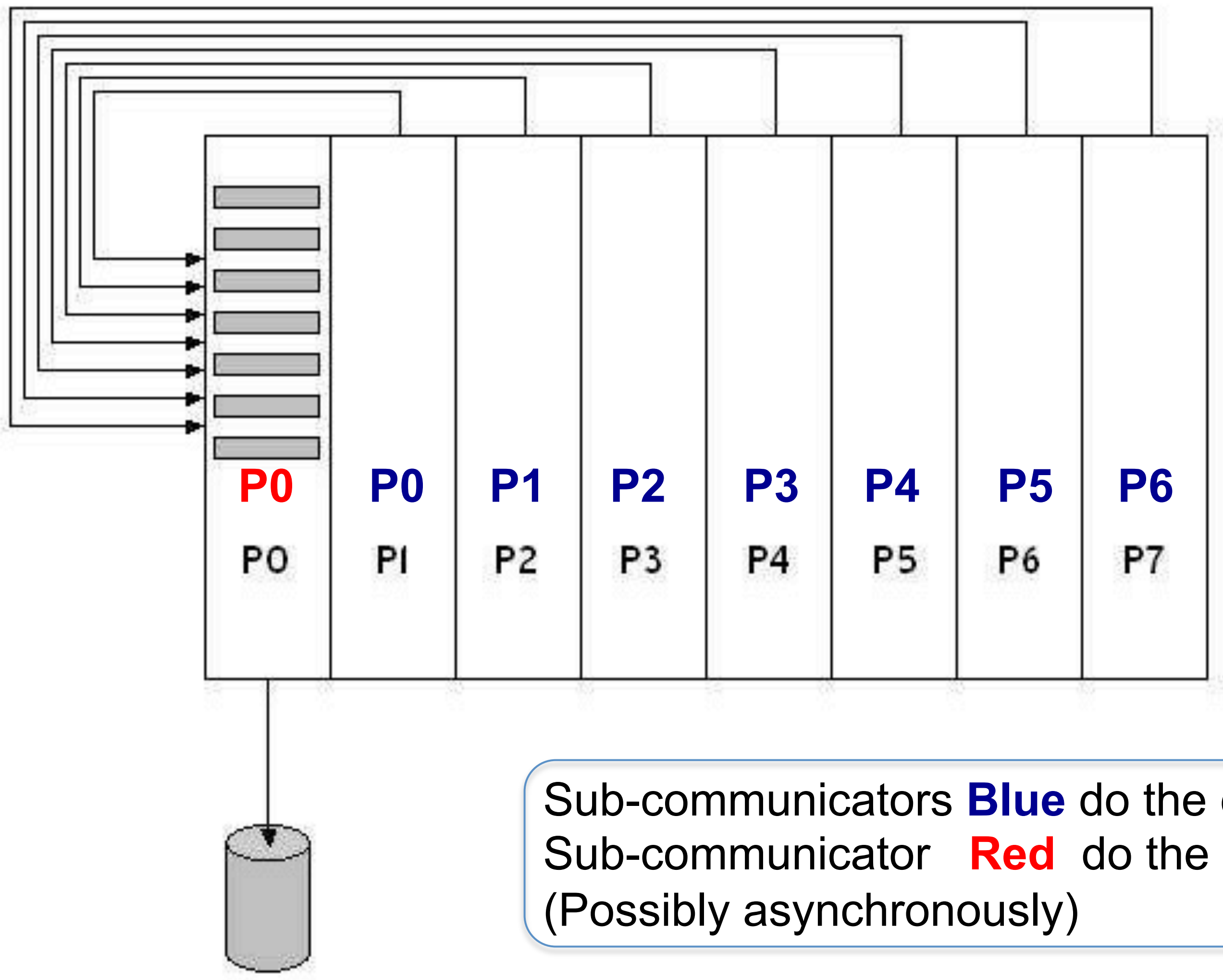
The problem of the I/O



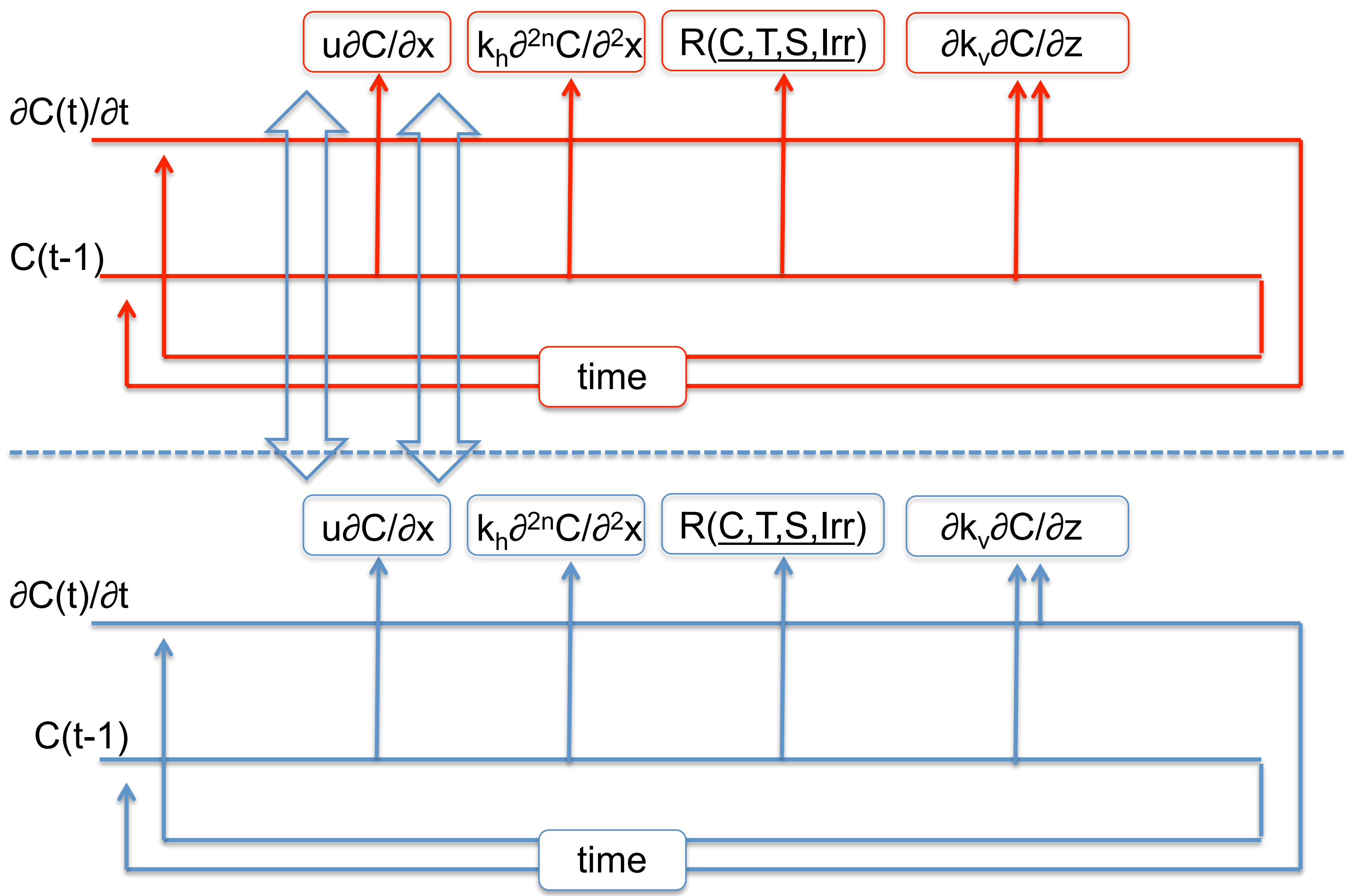


When P0 does I/O the others wait

Using different communicators



Sub-communicators **Blue** do the calclulation
Sub-communicator **Red** do the I/O
(Possibly asynchronously)



u∂C/∂x
$k_h \partial^{2n}C/\partial^2 x$
R(C,T,S,Irr)
$\partial k_v \partial C/\partial z$
∂C(t)/∂t
C(t-1)
time
u∂C/∂x
$k_h \partial^{2n}C/\partial^2 x$
R(C,T,S,Irr)
$\partial k_v \partial C/\partial z$
∂C(t)/∂t
C(t-1)
time

Thank you for the attention!