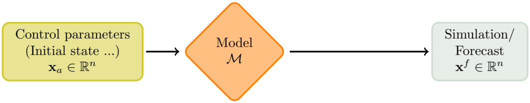
NEMOVAR, a data assimilation framework for NEMO

Arthur Vidard

September, 19$^{\mathrm{th}}$ 2008

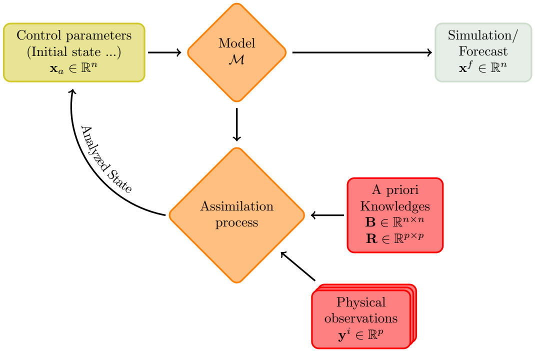
Data Assimilation

- Forecast is produced by integration of a model from an initial state
- Data Assimilation combines in a coherent manner all the available informations to retrieve an optimal initial state and then predict it:
    - Mathematical information : model
    - In-situ and remote measurements of the true state
    - A priori knowledges (background, errors . . .)



Data Assimilation

- Forecast is produced by integration of a model from an initial state
- Data Assimilation combines in a coherent manner all the available informations to retrieve an optimal initial state and then predict it:
  - Mathematical information : model
  - In-situ and remote measurements of the true state
  - A priori knowledges (background, errors . . .)



Variationnal Data Assimilation : The Optimality System

The state of the flow is governed by a partial differential system

$$\begin{cases} \dfrac{\partial \mathbf{x}}{\partial t} & = & \mathcal{M}(\mathbf{x}, U) \\ \mathbf{x}(0) & = & V \end{cases}$$

$U = U(x, y, z, t)$: unknown parameters
$V = V(x, y, z)$: initial condition

Cost function

$$J(U,V) = \frac{1}{2}\int_0^T \|\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}\|^2_{\mathcal{O}} + \frac{1}{2}\|U - U_0\|^2_{\mathcal{U}} + \frac{1}{2}\|V - V_0\|^2_{\mathcal{V}}$$

Adjoint (backward) model

$$\begin{cases} \dfrac{\partial \mathbf{p}}{\partial t} + \left[\dfrac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T . \mathbf{p} & = & \left[\dfrac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]^T . (\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}) \\ \mathbf{p}(T) & = & 0 \end{cases}$$

Gradient of the cost function

$$\nabla J = (\nabla_U J, \nabla_V J)^T = \left(-\left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T . \mathbf{p}, -\mathbf{p}(0)\right)^T$$

Variationnal Data Assimilation : The Optimality System

The state of the flow is governed by a partial differential system

$$\begin{cases} \dfrac{\partial \mathbf{x}}{\partial t} = \mathcal{M}(\mathbf{x}, U) \\ \mathbf{x}(0) = V \end{cases}$$

$U = U(x, y, z, t)$: unknown parameters
$V = V(x, y, z)$: initial condition

Cost function

$$J(U, V) = \frac{1}{2}\int_0^T \|\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}\|^2_{\mathcal{O}} + \frac{1}{2}\|U - U_0\|^2_{\mathcal{U}} + \frac{1}{2}\|V - V_0\|^2_{\mathcal{V}}$$

Adjoint (backward) model

$$\begin{cases} \dfrac{\partial \mathbf{p}}{\partial t} + \left[\dfrac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T.\mathbf{p} = \left[\dfrac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]^T.(\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}) \\ \mathbf{p}(T) = 0 \end{cases}$$

Gradient of the cost function

$$\nabla J = (\nabla_U J, \nabla_V J)^T = \left(-\left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T.\mathbf{p}, -\mathbf{p}(0)\right)^T$$

Variationnal Data Assimilation : The Optimality System

The state of the flow is governed by a partial differential system

$$\begin{cases} \dfrac{\partial \mathbf{x}}{\partial t} = \mathcal{M}(\mathbf{x}, U) \\ \mathbf{x}(0) = V \end{cases} \qquad \begin{array}{ll} U = U(x,y,z,t) & : \text{unknown parameters} \\ V = V(x,y,z) & : \text{initial condition} \end{array}$$

Cost function

$$J(U,V) = \frac{1}{2}\int_0^T \|\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}\|_{\mathcal{O}}^2 + \frac{1}{2}\|U - U_0\|_{\mathcal{U}}^2 + \frac{1}{2}\|V - V_0\|_{\mathcal{V}}^2$$

Adjoint (backward) model

$$\begin{cases} \dfrac{\partial \mathbf{p}}{\partial t} + \left[\dfrac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T . \mathbf{p} = \left[\dfrac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]^T .(\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}) \\ \mathbf{p}(T) = 0 \end{cases}$$

Gradient of the cost function

$$\nabla J = (\nabla_U J, \nabla_V J)^T = \left(-\left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T . \mathbf{p}, -\mathbf{p}(0)\right)^T$$

Variationnal Data Assimilation : The Optimality System

The state of the flow is governed by a partial differential system

$$\begin{cases} \dfrac{\partial \mathbf{x}}{\partial t} = \mathcal{M}(\mathbf{x}, U) \\ \mathbf{x}(0) = V \end{cases}$$

$U = U(x, y, z, t)$: unknown parameters
$V = V(x, y, z)$: initial condition

Cost function

$$J(U, V) = \frac{1}{2}\int_0^T \|\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}\|^2_{\mathcal{O}} + \frac{1}{2}\|U - U_0\|^2_{\mathcal{U}} + \frac{1}{2}\|V - V_0\|^2_{\mathcal{V}}$$

Adjoint (backward) model

$$\begin{cases} \dfrac{\partial \mathbf{p}}{\partial t} + \left[\dfrac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T . \mathbf{p} = \left[\dfrac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]^T . (\mathcal{H}[\mathbf{x}] - \mathbf{y}_{obs}) \\ \mathbf{p}(T) = 0 \end{cases}$$

Gradient of the cost function

$$\nabla J = (\nabla_U J, \nabla_V J)^T = \left(-\left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T . \mathbf{p}, -\mathbf{p}(0) \right)^T$$

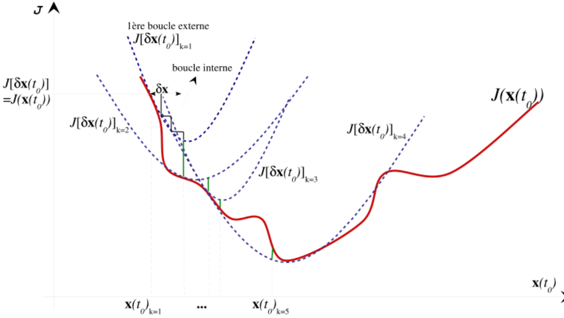
Incremental formulation

$\mathcal{H}$ and $\mathcal{M}$ are linerarized around $\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_0)$
The minimisation problem is approximated with a sequence of convex problem:

$$J(\delta \mathbf{x}_0) = \sum_{i=0}^{N} \|\mathbf{d}^i - \mathbf{H}[\mathbf{M}_i(\delta \mathbf{x}_0)]\|^2_{\mathcal{O}} + \|\delta \mathbf{x}_0\|^2_{\mathcal{X}}$$

with $\mathbf{d}_i = \mathbf{y}_i - \mathcal{H}[\mathcal{M}_i(\mathbf{x}_0)]$ and $\mathbf{M} = \frac{d\mathcal{M}}{dx}$ $\mathbf{H} = \frac{d\mathcal{H}}{dx}$

- Easier minimisation
- Allow the use of simplified physics / degraded resolution for $\mathbf{M}$ and $\mathbf{M}^T$
- Has to fulfil the Tangent-Linear Hypothesis $\mathcal{M}_i(\mathbf{x}_0 + \delta \mathbf{x}_0) \approx \mathcal{M}_i(\mathbf{x}_0) + \mathbf{M}_i(\delta \mathbf{x}_0)$



(Nucleus for European Modelling of the Ocean)

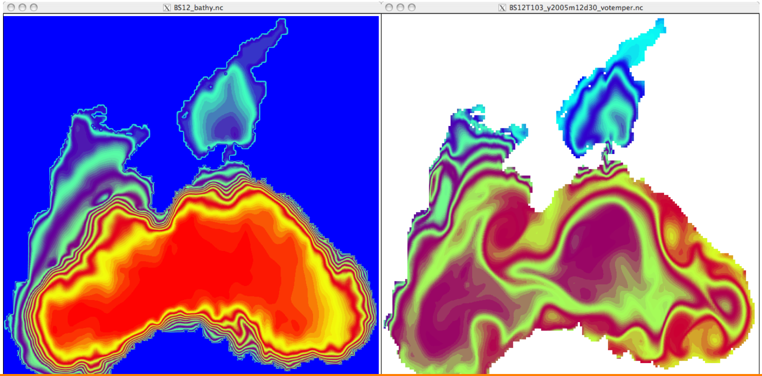
is a state-of-the-art modeling framework for oceanographic research and operational oceanography. This framework is intended to be interfaced with the remaining component of the earth system (atmosphere, land surfaces, ...). NEMO is distributed under CeCILL license. The major partners are organized within the NEMO Consortium, including CNRS, Mercator-Ocean, UKMO and NERC.
NEMO includes three engines (or components):

- OPA9 - the new version of the OPA ocean model.
- LIM2 - the new version of the Louvain-la-Neuve sea-ice model .
- TOP1 - a transport component based on OPA9 tracer advection-diffusion equation (TRP) and a biogeochemistry model which include two components: LOBSTER and PISCES.

http://www.lodyc.jussieu.fr/NEMO/
It is widely used by the scientific community used operationally by Mercator Ocean and soon by the UK-MetOffice and the ECMWF.

Black Sea configuration

- Based on the Mercator-Ocean global 12th of degree configuration (more informations on `www.mercator-ocean.fr/html/science/piste_rouge/model/model_orca12_en.html`
- Primitive equation model, Model state variables : temperature, salinity, zonal and meridian velocity, free surface elevation
- The BS12 configuration contains $174 \times 170 \times 42$ grid points ($174 \times 130 \times 42$ without the Azov Sea). Requires $\approx$ 820 MB ($\approx$ 620 MB)
- Can make use of the AGRIF nesting facilities (`ljk.imag.fr/MOISE/AGRIF`)
- and the NEMOVAR framework



The NEMOVAR Project

NEMOVAR was initiated in late 2005 and got support from LEFE/ASSIM (INSU), GMMC (Mercator) and TOSCA (CNES) and more recently ANR.

It gather researchers from INRIA, CERFACS, ECMWF, LOCEAN, LPO, LEGI, UK-MetOffice and Mercator-Ocean.

The aim of this project is to develop a multi-incremental variational data assimilation for the ocean model OPA9/NEMO for operational application and research.

Technical requirement:

- Efficient (parallel computing, memory management).
- Flexible (adapted to different kind of configuration from large scale global to high resolution local)
- Portable (script and source code)

Multi-incremental

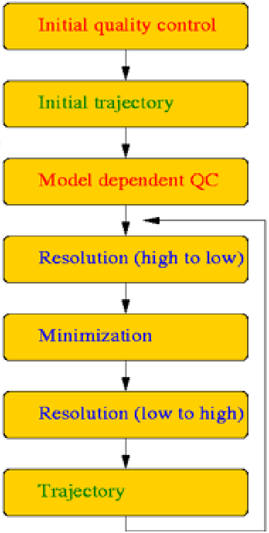
Planned data flow illustrated on the right:

- Initial quality control performs basic tests on the data.
- The initial trajectory computes the observation misfit and background state.
- Observation misfit and other information to perform an additional QC.
- The inner loop consist of
  - Change of resolution (high to low)
  - Minimization (3D-VAR or 4D-VAR)
  - Change of resolution (low to high)
- Update trajectory

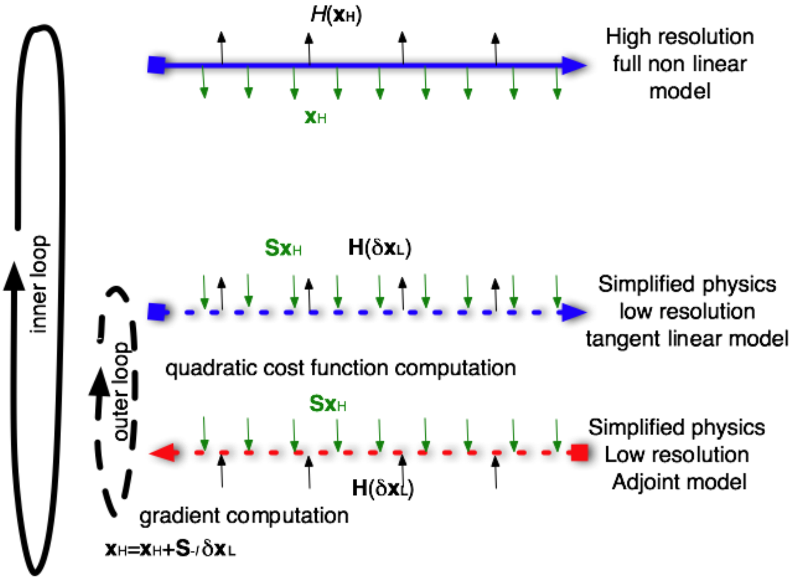
Multiple inner loops can be used

The same data flow can be used for other non-variational data assimilation schemes.

Common framework for assimilation with NEMO?



Multi-incremental



Current status

NEMOVAR

A 3D-FGAT *i.e.* is now available (An incremental variational data assimilation scheme where $\mathbf{M}$ and $\mathbf{M}^T$ are approximated by I).

It can assimilate in situ Temperature and Salinity, SSH (maps and along track), SST, SSS and current velocities.

This system was validated on a 45+ years reanalysis of the global ocean at 1° resolution. It fulfilled the scientific and technical requirements.

NEMOTAM

We tried to use the automatic differentiation tool TAPENADE (projet TROPICS, INRIA Sophia) and manage to get a working tangent and adjoint of a given configuration.

But the handling of the non linear trajectory was severely incompatible with the multi-incremental approach. Moreover the computing efficiency was not satisfactory. This work will continue within the framework of the ANR project VODA (MOISE / TROPICS / NEMO-Team)

But for the short term it is Hand coding time.
First version expected by the end of 2008.

In ADAMS

In the proposal

- Set up the Black Sea configuration BS12 properly.
- Control of Model Error in VDA (with Y. Trémolet)
- Assimilation of Lagrangian observation (with M. Nodet)

Other possibilities

- Direct (or not) assimilation of images
- Sensitivity analysis
- ...