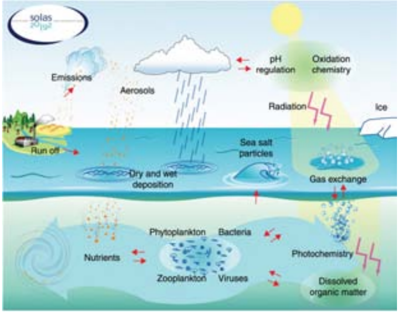
Data assimilation in 3D ocean coupled physical-bigeochemical models: state and parameter estimation, using nonlinear extensions of the Kalman filter

Maeva Doron, David Béal, Clément Fontana, Pierre Brasseur and Jean-Michel Brankart

Laboratoire des Ecoulements Géophysiques et Industriels (LEGI-MEOM team), Grenoble, France

doron@hmg.inpg.fr

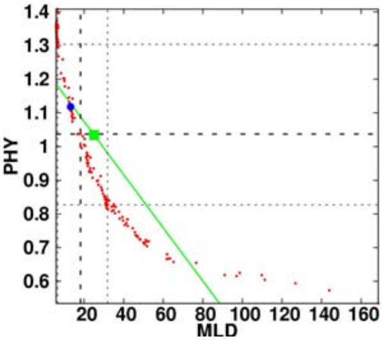
Why use 3D ocean coupled physical-biogeochemical models?

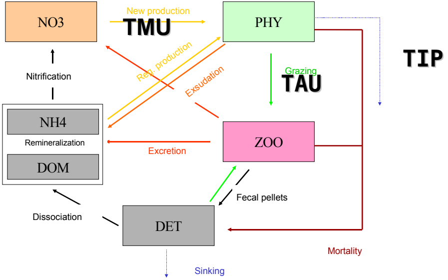


- **Coupled physical-biogeochemical models** deal with the biogeochemical state of the oceans

- **Phytoplankton** is the link between inorganic and organic matter
  - can **grow** when there is availability of light and nutrients
  - is the **first trophic level** in the food chain

Why non-linear and non-Gaussian extensions are needed?

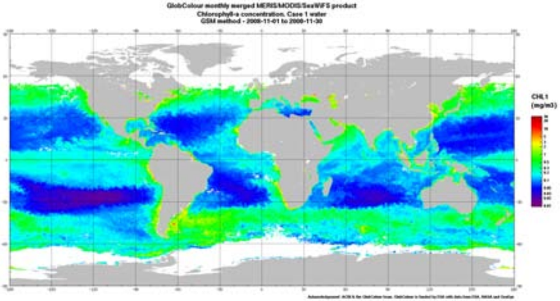




Ensemble simulations with a **3D coupled physical-biogeochemical model**

- **non-Gaussian** behavior
- **non-linear** dependence between key variables

Hence, **linearized** methods, like the Kalman filter, are **far from optimal**



- **Satellite ocean colour** is a unique way of observing biology in the oceans (chlorophyll *a*)
- **Wealth** of observations: global scale, fine resolution, 10-year-archive, **but**: surface measurements, cloud coverage, poorly known observation error

Context

Strong interest in ocean coupled physical-biogeochemical models

- Tool to quantify fluxes in carbon cycle
- Estimation and forecast of biological state of the ocean by operational oceanography centers

Difficulties in using coupled models

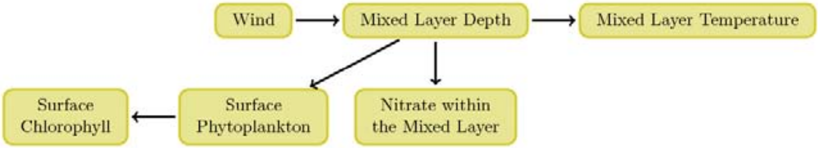
- Very sensitive to the physical forcing
- Biogeochemical laws and parameters are uncertain

Satellite ocean colour sensors provide chlorophyll *a* concentrations

- Wealth of observations
- But only surface data

Methodology developped to **reduce the uncertainty in the coupled models**, such as the initial condition, the parameterisation or the physical forcing with data assimilation of **ocean colour** data using sequential stochastic methods.

Cascade of errors in the coupled model:



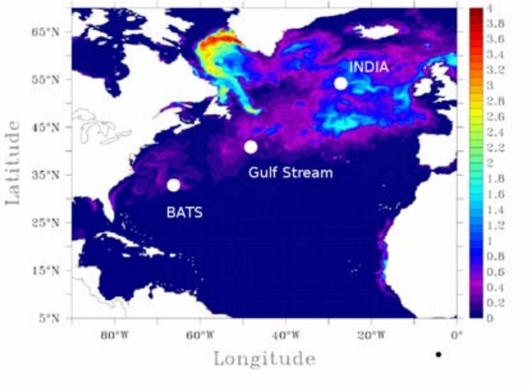
Protocol to generate the Gaussian perturbations

- ERA40 wind forcing database (1985-2000)
- Covariance of the wind variability computed for the spring season
- 50 EOF kept (80% of the variance)
- Wind perturbation randomly sampled with this covariance

Gaussian wind perturbation: **distribution of the output?**

The coupled physical-biogeochemical model response to the wind perturbation

- Experiments last 1 month from April 16, 1998
- 200 ensemble members

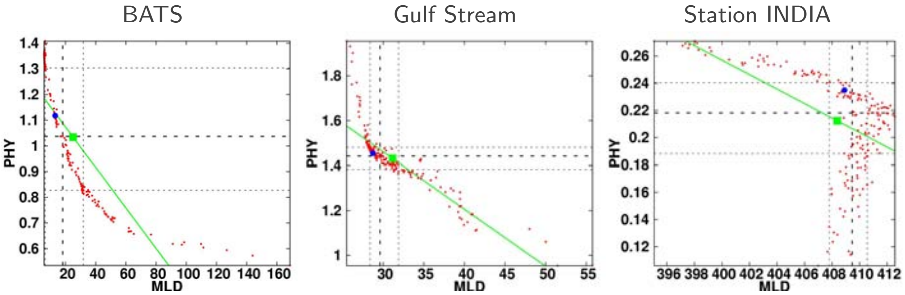


Standard Deviation of the ensemble for surface phytoplancton after 1 day

Very inhomogeneous response in phytoplankton

Non-linear response of the model

Outputs from the 200 runs after 1 day for three stations

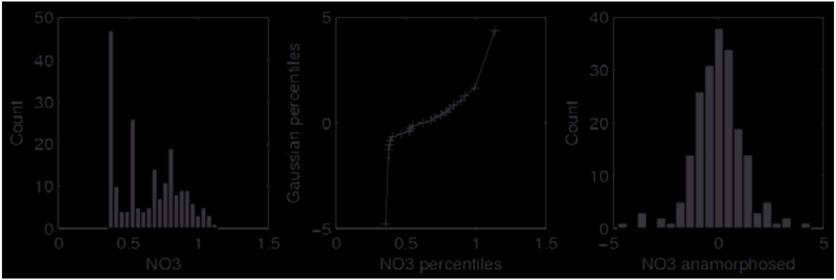


- **Non-linear** relationship between PHY and MLD
- For the BATS station, the linear approximation could be valid for smaller range of perturbation
- **Very different** behavior in different regions

How to cope with non-Gaussian distributions?

Example of the nitrate concentration (NO3 in $mmolNm^{-3}$) after 1 day in the Gulf Stream station

- Anamorphosis first introduced by Bertino et al. (2003)
- Here, **piecewise linear** change, mapping the ensemble percentiles to **Gaussian percentiles**.

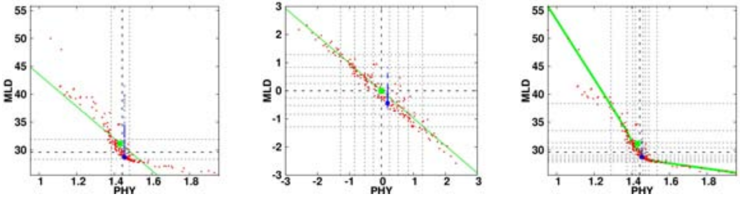


- The anamorphosis is done **locally and independantly for each variable**

Example of observational update at one station

After 1 day, here in Gulf Stream station

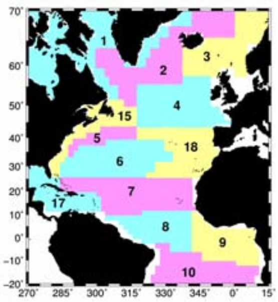
- Perfect observation of PHY, update of Mixed Layer Depth

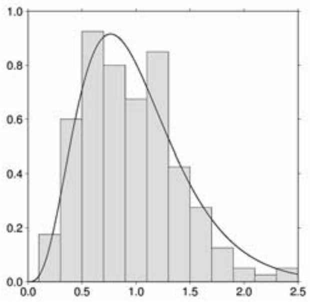


- **Left**: Analysis in the initial space using linear regression line with the original ensemble
- **Right**: Analysis in the transformed space using linear regression line with the transformed ensemble
- Non-linear relationship between the two variables better taken into account
- **Reduction in the dispersion of the analysed values**

D. Béal, P. Brasseur, J.-M. Brankart, Y. Ourmières and J. Verron, *Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis, Ocean Science, 2010*

The ecological provinces



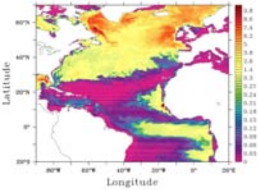


- Ecological provinces defined by Longhurst (1995, 2007)
- Assumption that three key biogeochemical parameters are independent on each region
- The rates have positive values: Gamma distribution are tailored such that the average is the reference value, the 5% and 95% percentiles are 0.5 and 2 times the reference value (validated simulation).

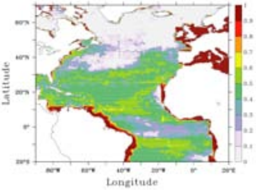
Monte Carlo experiment

- Monte Carlo experiment: 200 simulations with perturbed parameters maps from the same identical initial conditions and same forcing
- The only difference between simulation is the parameter values, imposed the 15^{th} of March 1998 and held constant
- The simulation length is 1 month

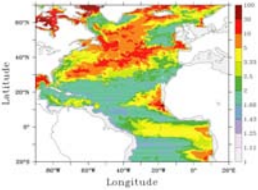
Std dev. after 30 days



5% percentile/Median



95% percentile/Median

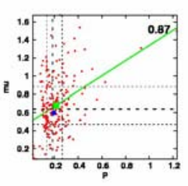


- Large spatial heterogeneity in the model response

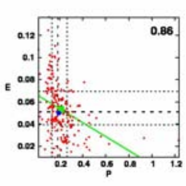
Model response in region 18, North Atlantic Subtropical Gyre (36°N 19°W)

- After 30 days, surface values of the 200 simulations are extracted
- Blue is for the reference, red points stand for the 200 members and green is the ensemble average

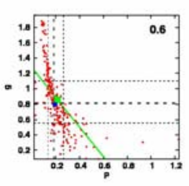
PHY-TMU: 0.37



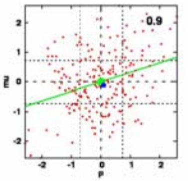
PHY-TIP: -0.37



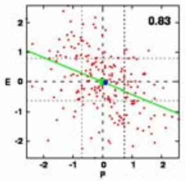
PHY-TAU: -0.63



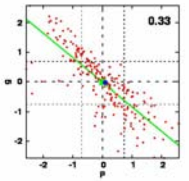
PHY TMU anam: 0.32



PHY-TIP anam: -0.41



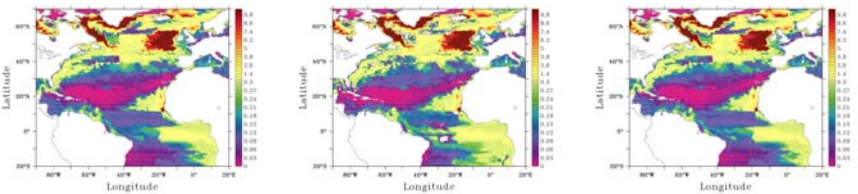
PHY-TAU anam: -0.82



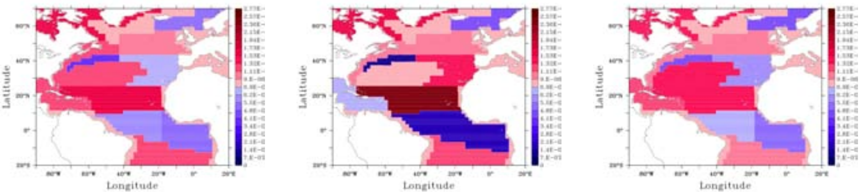
Set up of the twin experiments

- The **false** ocean is the **reference** simulation (Ourmières et al. JMS 2009) in which the parameters are uniform
- The **true** ocean is an **independent** simulation (perturbed on the same assumptions than for the Monte Carlo experiment)
- $\hat{P}^f$ is approximated by the multivariate EOFs done on the 200-member ensemble
- R : the observation error is chosen very low
- The analysis is either done in a classical way (linear) or performed with the anamorphosed variables (nonlinear)

observed PHY | linear analysed PHY | nonlinear analysed PHY



true TMU | linear analysed TMU | nonlinear analysed TMU

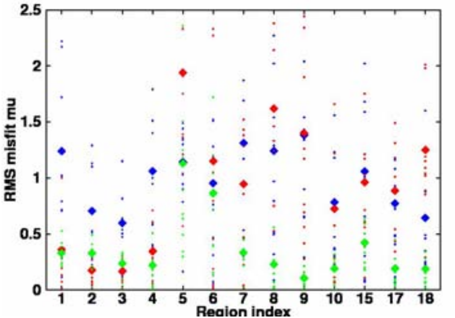


- The linear analysis produces artefacts: negative values in P and parameters
- The nonlinear analysis is more efficient in retriving the parameters

Objective norm: into the anamorphosed space

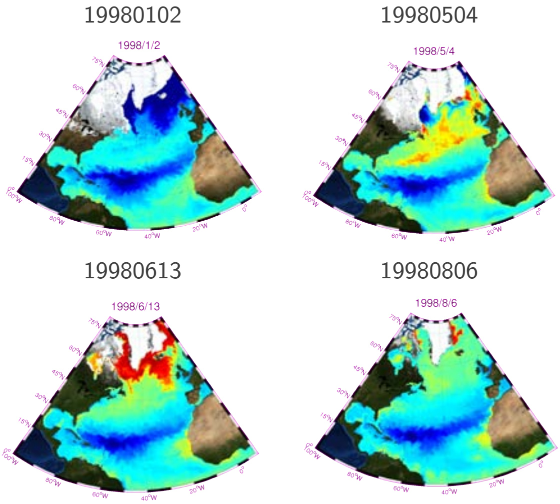
- All RMS are calculated in the anamorphosed space: variables and paremeters roughly between -3 and 3
- Statistics on RMS for 10 different observations from 10 true oceans: mean and STD of the RMS (similar results for the two other parameters)

RMS for TMU



M. Doron, P. Brasseur and J.-M. Brankart, *Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical-biogeochemical model : twin experiments, in minor revision for Journal of Marine Systems, 2011*

Free simulation in the North Atlantic during the spring bloom

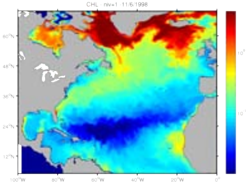


- The development of the bloom can be seen in the free simulation

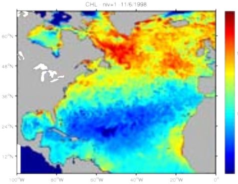
Implementation of the SEEK filter

- The SEEK filter (Pham et al. 1998, Brasseur and Verron 2006) was implemented
- Use of the software SESAM (http://www-meom.hmg.inpg.fr/Web/Outils/SESAM/sesam.html)
- EOFs for the current season (kept on 2 months), calculated on model outputs from the free run (every 2 days)
- Observation error taken equal to 30% of the chlorophyll a concentration
- Ocean color data from SeaWiFS-GAC (last reprocessing): with observations averaged every 4 days

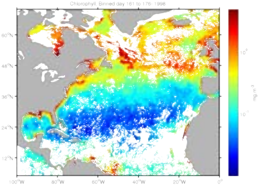
FREE



SEEK



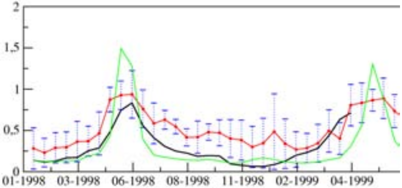
DATA



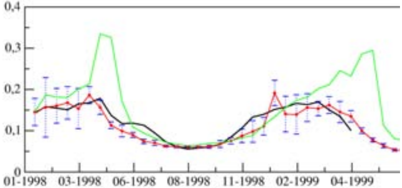
- Average on the first 15 days of June 1998
- SEEK has a positive impact of the estimation of the bloom :reduction of the very large concentrations at high latitudes

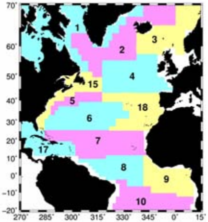
Diagnostic on different regions

Region 4

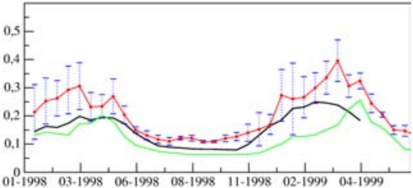


Region 6



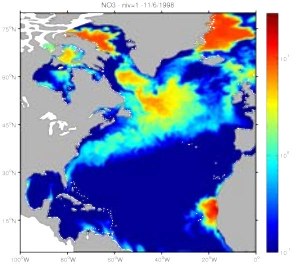


Region 18

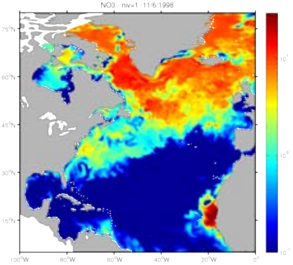


- SEEK performs well in oceanic regions and not so well in coastal regions (not shown here)

Nitrates FREE



Nitrates SEEK



- SEEK allows to correct for non observed variables such as nitrates, which is beneficial
- Nitrates are poorly sampled in the world ocean
- This data assimilation system could provide better climatologies of an **unobserved variable**

Summary - Conclusions

- Model response to **wind forcing and parameter uncertainty** tested with a Monte Carlo method: **nonlinear** response and **non-Gaussian** distributions
- **Anamorphosis**: the implementation of a **nonlinear transformation** allows to get closer to linearity
- **Parameter estimation** in a coupled model is possible from P observations (twin experiments) using anamorphosis
- **Data assimilation with SEEK** and ocean color data provides good results on observed and unobserved variables

Next steps

- Use of **real ocean color data for parameter estimation**
- Extension of the **state estimation experiment with SEEK** on longer timescales and implementation of anamorphosis to consider its impact
- Transfer of the **anamorphosis** methodology to other fields of application of data assimilation, for instance with altimetry data, in cases where the model show nonlinear or non-Gaussian behavior (J.-M. Brankart et al., in preparation)