Published in: The Mediterranean Sea (2005), pp. 361-385
Status: Postprint (Author's version)

Applicability of Models in the Field of Environmental Chemistry of the Mediterranean Sea

J.M. Beckers

MARE-GHER, University of Liège, Sart-Tilman B5, 4000 Liège, Belgium

Research Associate, National Fund for Scientific Research, Belgium. This is MARE publication MARE057.

Abstract: The present chapter highlights the problems facing when modelling the marine environment, in particular for biochemical aspects. It is shown how the marine environment differs from other fluids and how this affects modelling. General modelling approaches are shown and the problem of the vaste range of scales encountered is adressed. In addition to mathematical modelling problems, practical questions of adequate boundary and initial conditions are then further analysed with respect to biochemical components carried by the flow. Finally questions of errors, validation and data assimilation encountered during the analysis of the model's quality are adressed.

1. Introduction

Before looking at the modelling aspects of the marine environment and in particular its biochemistry, we should show the particularities of the system compared to those of other systems (as laboratory experiments) being modelled currently. When working in marine systems, we face the following special problems which must be taken into account:

• From a dynamical point of view:

- The system is characterised by a very large spectrum of scales and processes. Scales range from millimetres (and fractions of seconds) to planetary scales (and centuries) not only for physical processes (from turbulence to global circulation, see Fig. 3), but also for biological or chemical processes.
- The aspect ratio for large-scale processes is extremely small, meaning that the vertical scales are much smaller than the horizontal scales.
- The system under study is generally stratified, with the combination of currents and stratification controlling the exchanges between the surface layers and deeper layers.
- For larger scales, hydrodynamics must take into account earth rotation. In combination with stratification, fronts can be maintained by the Coriolis force, which then controls cross-frontal exchanges.
- Water movements are guided by topography and land boundaries, and biochemical exchanges at the land-sea interface also influence the marine system.

• From an observational and "experimental" point of view:

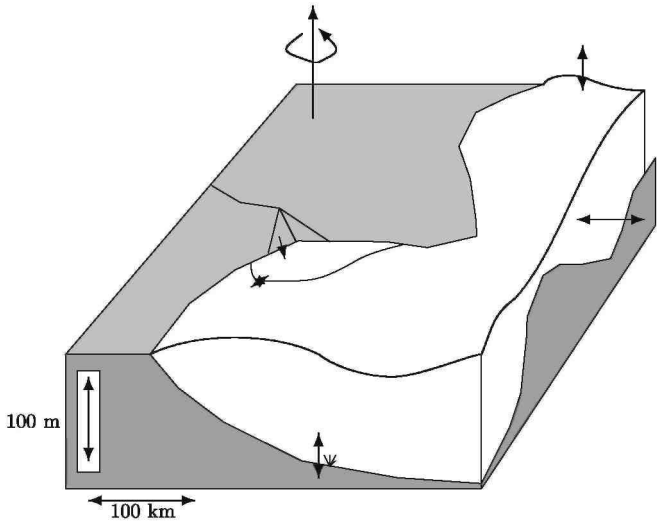
- Data available to calibrate, validate, initialise and force models are generally incomplete due to logistic constraints.
- Experiments in the natural environment can rarely be repeated or controlled as nature evolves due to forcings, which we are generally not able to modify (e.g. wind stress, heat fluxes, most river discharges).
- Forcings at air-sea boundaries or land-sea boundaries are not controllable, and they are often not known very well either, which limits the precision of model simulations. The most imprecise conditions are generally those related to the artificial boundaries used to isolate the system under study from the rest of the marine system (e.g. during a study of the Gulf of Cadiz one does not intend to analyse and model the Atlantic Ocean, which of course influences the evolution of the Gulf of Cadiz).

• From a political and socioeconomic point of view:

- The system under study is influenced by adjacent systems and in reality is often interacting with and modifying the other system in turn. Remote influences can lead to problems which are out of control for local factors only able to manage the system within their sphere of influence.

All these exchanges (schematised in Fig. 1) and particularities explain why it is generally not possible to simply use a "black box" model developed for laboratory experiments or industrial application simulations to model marine systems. Adaptations need to be made to be able to deal with the particularities mentioned above. In the following, we will show how such adaptations are carried out.

Fig. 1: *Schematic exchanges between the sea and the surrounding environments (land, air, rivers, sediments, open ocean)*



2. Modelling Aspects

When developing a model, the first characteristic one has to define is its objective. Among the objectives of a model, one can mention:

• Models aimed at improving the understanding of the functioning of a system (e.g. process models).
• Models aimed at quantifying processes understood theoretically but not with real precision (e.g. flux estimates).
• Models aimed at predicting evolution of scenarios (e.g. estimation of impacts of changes in river pollution on a coastal ecosystem).
• Models aimed at daily use for management purposes (e.g. managing allowed waste water discharges according to prescribed maximum levels of concentrations).

Such types of objectives are frequently found, be it for hydrodynamics, bio-geochemical or socio-economic aspects. The choice of an appropriate model depends on the objective at hand and should be done with care. In particular, the model should contain the necessary and sufficient parameters to be calculated that allow one to characterise the system and answer the question that the model is used for.

The mathematical model developed, the numerical methods used and the data required vary considerably from one case to another according to the objective (e.g. for operational purposes a robust method with regular data input is essential, whereas for theoretical studies more technologically advanced models with fewer data requirements can be used). Models, be it the mathematical version or its numerical implementation, thus differ largely depending on their objectives and, as we will see later, scales.

There is, however, one thing all deterministic models have in common: the way the models are constructed. They always involve some way of:

• Expressing budgets of

- Mass of seawater
- Momentum (Newton's law to derive the current fields)
- Energy (giving access to temperature evolution)
- Biogeochemical components (salinity, pollutants, tracers...)

• To which one adds

- Constitutive equations (e.g. diffusion laws)
- State equations (e.g. ocean water state equation)
- Interaction laws between components (e.g. chemical reactions, phytoplankton uptake law)

These ingredients are used to elaborate a mathematical model by making budgets on control volumes V (Fig. 2).

If the control volume is infinitely small (i.e. in practice much smaller than the scales of interest), one can derive the mathematical formulations of the time-space evolution of the system.

An example of a budget for the concentration of constituent c^a over a control volume in a flow of velocity v (Fig. 2) leads to the following time (f) evolution law:

$$\frac{\partial c^a}{\partial t} + \nabla \cdot (c^a \boldsymbol{v}) = Q^a - \nabla \cdot (c^a \boldsymbol{m}^a) + \nabla \cdot (\lambda^a \nabla c^a) \qquad (1)$$

where Q^a is the local source of the constituent (through chemical reactions for example), m^a a migration/sedimentation velocity, λ^a a diffusion coefficient (generally turbulent) and $\nabla = \boldsymbol{e}_x \partial/\partial x + \boldsymbol{e}_y \partial/\partial y + \boldsymbol{e}_z \partial/\partial z$ the classic derivation operator.

Applying this approach to the salinity, temperature, momentum, mass and concentration of any biochemical tracer, one can develop the so-called primitive equation models (Table 1), which describe in principle the fully 3D time-dependent structure of the ocean, including its biochemical components.

From a mathematical point of view, adding appropriate initial and boundary conditions to these equations is sufficient to predict the evolution of the system.

Fig. 2: *The time evolution of the total mass of a given constituent c^a in a fixed volume V of surface S is equal to the sum of local production (destruction) ρQ^a of the constituent c^a within the volume V and the transport of the constituent across the surface S. This transport consists of a general transport by the fluid with velocity v and a movement of the constituent relative to the fluid. These relative movements are related to diffusive processes d^a and organised movements such as migration/sedimentation m^a*

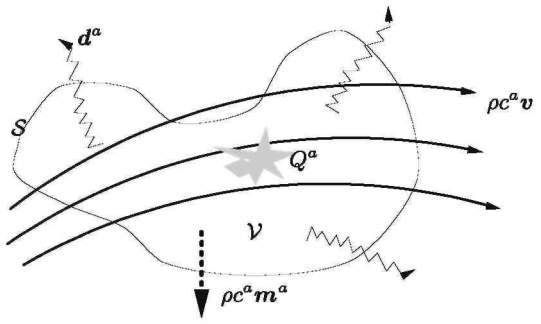


Table 1: *Primitive equation model for velocity v, temperature T, salinity S, biochemical component à and state equation for density ρ of reference value ρ_0. g is gravity, $\boldsymbol{\Omega}$ the earth rotation vector, Q^e the radiative heat source and p pressure*

$$\nabla \cdot \boldsymbol{v} = 0 \qquad (2)$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \nabla \cdot (\boldsymbol{v}\boldsymbol{v}) + 2\boldsymbol{\Omega} \Lambda \boldsymbol{v} = -\nabla q + \boldsymbol{b} + \nabla \cdot \left(\nu \nabla \boldsymbol{v}\right) \qquad (3)$$

$$\frac{\partial T}{\partial t} + \nabla \cdot (\boldsymbol{v}T) - \frac{\alpha T}{\rho_0 c_p} \frac{\mathrm{d}p}{\mathrm{d}t} = \nabla \cdot (\lambda^T \nabla T) + \frac{Q^e}{c_p} \qquad (4)$$

$$\frac{\partial S}{\partial t} + \nabla \cdot (\boldsymbol{v}S) = \nabla \cdot (\lambda^S \nabla S) \qquad (5)$$

$$\frac{\partial c^a}{\partial t} + \nabla \cdot (c^a \boldsymbol{v}) = Q^a - \nabla \cdot (c^a \boldsymbol{m}^a) + \nabla \cdot (\lambda^a \nabla c^a) \qquad (6)$$

$$\rho = \rho(T, S, p, \ldots), \quad b = -\frac{\rho - \rho_0}{\rho_0} g, \quad q = \frac{p}{\rho_0} + gz \qquad (7)$$

Published in: The Mediterranean Sea (2005), pp. 361-385
Status: Postprint (Author's version)

3. Scale Dependence and Intermittency

In practice, modelling the system is not achievable without further analysing scales. Indeed, even if primitive equations were able to describe all scales of motion (Fig. 3), numerical constraints (grid and time step restrictions) only allow one to resolve scales down to a given resolution. Limitations in initial conditions and boundary conditions also lead to limitations in the timescale of prediction.

Therefore, basically one can only resolve part of the processes of the system, while the others are not predictable by the model. This does not mean, however, that the scales not resolved do not influence the scales that are modelled. It is clear that small-scale turbulent motion certainly influences the diffusion of tracers, even if the latter are modelled at much larger scales than the scale of the turbulence itself. In fact, we are faced with a problem of parameterisation, i.e. include the effect of unresolved processes in the model by empirical or statistical functions.

Unresolved processes can be of several types:

• Intermittent processes (e.g. flush discharges).
• Processes of spatial scales smaller than the grid size (e.g. mesoscale up-wellings at a typical 20-km scale, in Mediterranean models of 20-km resolution).
• Processes for which adequate forcing is not taken into account (e.g. diurnal thermocline evolution in models with seasonal sunlight evolution only).

The effect of the unresolved scales appears formally when averaging the equations so as to filter out the unresolved scales. When doing so, non-linearities in the equations will lead to terms which cannot be calculated without knowledge of the unresolved scales. This is a closure problem, and one has to find a way to assess the value of these terms in relation to the large-scale processes one is analysing. Basically, two approaches are used:

• Parameterisation based on experiments and theoretical work (this is done for the turbulent motions whose effects are modelled/parameterised by a diffusion law with a turbulent diffusion coefficient, Table 2).
• Explicit calculation by a specialised model of the smaller-scale processes. A typical application here is the use of nested models with higher resolution.

The fact that not all processes are resolved by models thus introduces some need for adaptation of the mathematical formulations.

Fig.3: *Scales and processes in the ocean [1]*

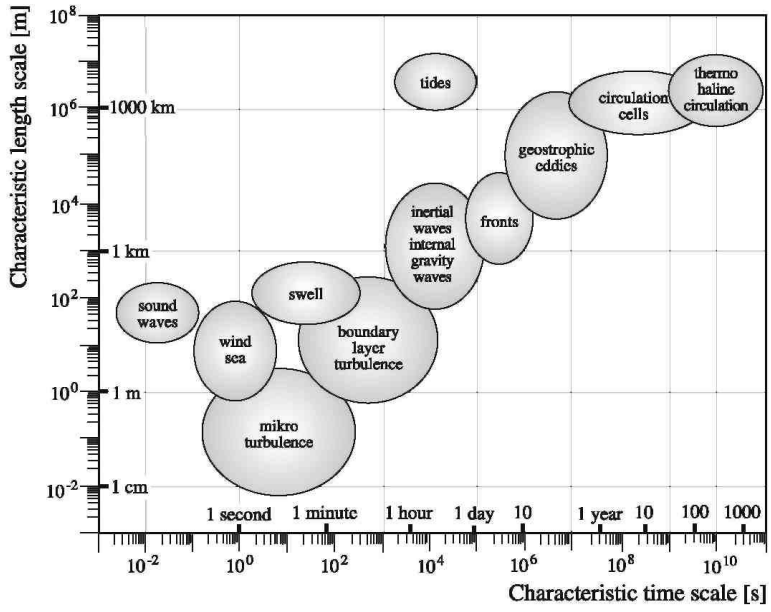


Table 2: *Turbulent closure scheme to calculate turbulent viscosity* $\tilde{\nu}$ *and turbulent diffusion coefficient* $\tilde{\lambda}$ *as a function of turbulent kinetic energy k, shear and stratification parameters, and a mixing length* l_0

$$\frac{\partial k}{\partial t} + \boldsymbol{u}\cdot\nabla k + u_3\frac{\partial k}{\partial x_3} = \tilde{\nu}M^2(1-R_\mathrm{f}) - \frac{k^2}{16\tilde{\nu}} + F_\mathrm{k} + \frac{\partial}{\partial x_3}\left[\tilde{\nu}\frac{\partial k}{\partial x_3}\right] \tag{8}$$

$$\tilde{\nu} = 1/2\sqrt{k}\,l_0(1-R_\mathrm{f}) \tag{9}$$

$$\tilde{\lambda} = \gamma\tilde{\nu}\sqrt{1-R_\mathrm{f}} \tag{10}$$

$$R_\mathrm{f} \equiv \frac{\tilde{\lambda}}{\tilde{\nu}}R_\mathrm{i}\,, \quad R_\mathrm{i} \equiv \frac{N^2}{M^2} \tag{11}$$

$$M^2 \equiv \left\|\frac{\partial \boldsymbol{u}}{\partial x_3}\right\|^2, \quad N^2 \equiv \frac{\partial b}{\partial x_3} \tag{12}$$

4. Numerical Model

Once the mathematical model is established including these parameterisations, a numerical method designed to provide efficient and accurate approximations to the solutions of the equations must be used. In particular, grid size, time steps and resolutions of the forcings must be coherent with the scales the mathematical model is supposed to resolve. (If the model includes parameterisation of turbulence but is supposed to resolve explicitly mesoscale processes, a resolution of less then the deformation radius must be used.) Also, the numerical techniques chosen must be able to accurately represent the dominant processes at stake. Since for biochemical models, very often advection is among the dominant processes, the advection scheme needs particular attention (it should be front preserving and not create negative concentrations, for example).

Typically, the ingredients for the GHER model [2] of the Mediterranean include:

- Finite-volume approaches on an Arakawa C grid (to conserve easily mass and concentrations).
- A *mode splitting* for the free surface (to follow free surface movements efficiently).
- A TVD advection scheme (to preserve fronts).
- A monotonic advection (to maintain concentrations positive).
- A generalised vertical coordinate (to follow the thermocline and bottom topography).
- Recursively bidirectional nesting (to allow for fine resolutions where needed).

The physical model and its numerical implementation are now relatively well mastered, but particular challenges appear when coupling biogeochemical models to the hydrodynamics.

5. Challenges in Coupling of Hydrodynamical Models with Ecosystem, Biochemical or Pollution Models

5.1 Scale Effects

In principle, one could simply adopt the same numerical and mathematical modelling approach as for the physical system. However, even if the physical solution at a given scale is the "real one", the biological system at this scale may be strongly influenced by unresolved processes, for example a sudden vertical mixing due to an unresolved wind event (Fig. 4) or a baroclinic instability and associated vertical velocities not resolved in a general circulation model. Parameterising these effects in strongly nonlinear biochemical models is quite delicate, and the parameterisations in such models are strongly dependent on the unresolved processes and therefore likely to be not very robust. (Even if a parameterisation was calibrated for a given system, if the intermittent process is delayed for some reason or differs from the statistical average, the biological system could react quite differently. For example, a delay in a storm event triggering a phytoplankton bloom is difficult to parameterise in a system were such storm events are only taken into account by statistical averages on the past.)

Similarly, the intermittency of riverine inputs (Figs. 5 and 6) also affects the system reaction and must be either taken into account explicitly or parameterised properly.

Fig.4: *Intermittent mixing and deepening of the thermocline [3] illustrated by shear of velocity M and Brunt-Väisälä frequency N as a function of depth and time*

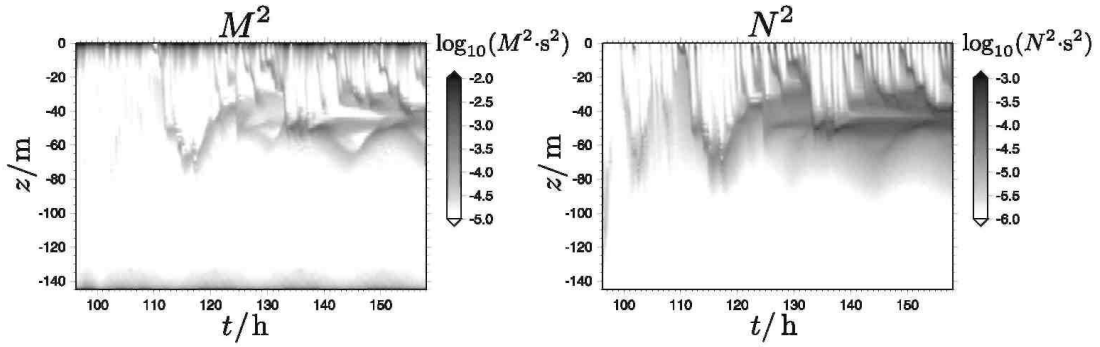
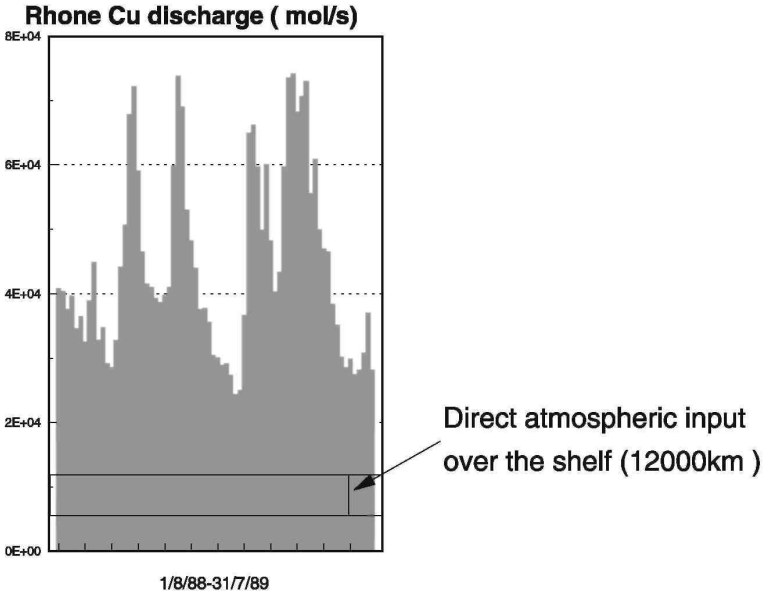


Fig. 5: *Intermittent discharge of Cu from the Rhone River into the Gulf of Lions as a function of time and comparison with atmospheric deposition estimates*



Open Boundaries

The open boundary problem is probably among the most crucial for regional models. Indeed, it is at this boundary that the information on the rest of the marine system is transferred between the regional model and the remaining system. Except in the case where nested models are used (in which case the larger scale models are presumed to model and predict the marine system beyond the regional model), assumptions on the system evolution at the arbitrary and artificial open-sea boundaries *must* be made. A particularly illustrative example [4] of the need for additional assumptions is the case of a model aimed at forecasting a river plume (of salinity lower than 17) by a box model (Fig. 7). Integrating the 3D equations over the domain of interest directly makes the fluxes across the boundaries of the domain appear, fluxes which depend on the internal structure of the plume (not known any more due to integration), the outside structure of the fields (not known by the limitation of the model domain), the velocity structures at the boundaries (not known unless predicted by another model) and the diffusion coefficients on the interface (which often depend on turbulence levels not resolved by a plume model).

Fig. 6: *Effect of changes in the wind pattern on Cu inflow on the shelf from the Liguro-Provencal current (left) and the deep waters (right) [5]*

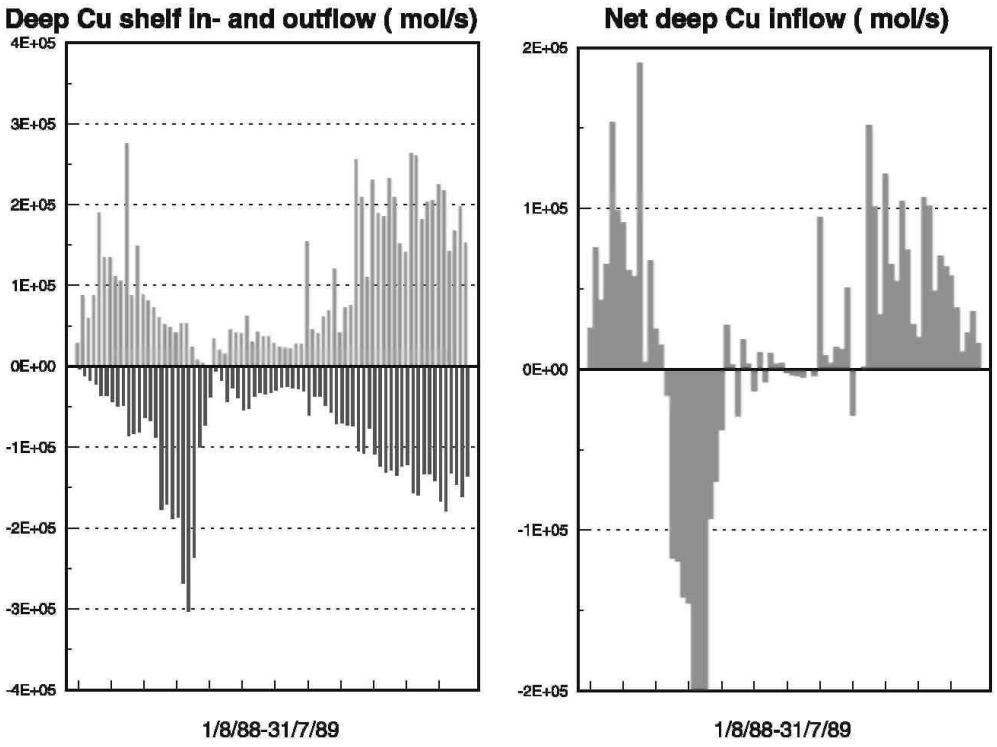
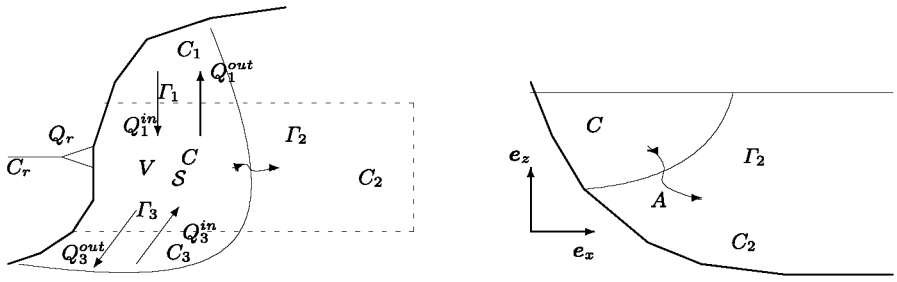


Fig. 7: *Schematic representation of the box model. V is the volume of water considered between fixed latitudes and a moving interface defined by S* = 17. *This volume is supposed not to detach from the river mouth, otherwise some lateral mixing exchanges should replace the river input. S is the sea surface delimited by the two latitudes and the intersection with the surface of constant salinity Γ2. Γ1 is the vertical surface defined by a given northern latitude and its intersection with the isosurface S* = 17. *This is the northern, upper limit of the box model.* Γ_2 *is the isosurface S* = 17 *delimited by the two latitudes and defines the eastern lateral limit of the box model.* Γ_3 *is the vertical surface defined by a given southern latitude and its intersection with the isosurface S* = 17. *This is the lower, southern limit of the box model.* C_r, C_1, C_2, C_3, C *are the concentrations of a tracer in the river, the northern part, the open sea, the southern part and the plume, respectively.*



The actual integration over the moving and open 0D box can be obtained mathematically, but due to these unknown structures needs additional hypotheses to close the formulation in terms of integrated variables only. One assumes, for example, that the total integral of source terms can be retrieved by applying the local source law P^C to the average concentrations. This is only true when interaction laws are linear or if the laws have been adapted (through a specific calibration) to represent large-scale interactions rather than local, physiologically based interactions. Similarly, the sea-surface interactions generally depend on the evolution of the constituent in the water mass itself. We must assume that the integral of the sea-surface exchange can be expressed in terms of the average concentration of the tracer. The same hypotheses apply to the fluxes across the fixed boundaries Γ_1 and Γ_3 (Fig. 7). Here one should bear in mind that these fluxes are dominated by advective fluxes whose

integration can be written as the product of the water mass flow and the concentrations of inflow or outflow.

Based on such assumptions, to be able to force such a box model one must at least have either field data or diagnosed values from the 3D model for the following parameters: the river discharge Q_r in m^3/s into the sea, the water inflow Q_1^{in} into the volume coming from the north, the outflow Q_1^{out} to the north on the northern boundary, the inflow from the south Q_3^{in}, the outflow to the south Q_3^{out}, the sea-surface flux counted positively upwards F^C, and the inflow concentrations in the north C_1 and the south C_3.

Providing such parameters are known, based on our assumptions and omitting notations for averages, the evolution of the average concentration in the box is governed by:

$$\frac{dC}{dt} = P^C + \frac{Q_r}{V}(C_r - C) + \frac{Q_1^{in}}{V}(C_1 - C) + \frac{Q_3^{in}}{V}(C_3 - C) - F^C\frac{S}{V} + \frac{A}{V}(C_2 - C) \quad (13)$$

where an additional equivalent diffusion coefficient A was defined to take into account unresolved exchanges. The calculation of the latter is based on an equivalent diffusion flux formulation for salinity or temperature. Its actual value for the total diffusion flux can be diagnosed thanks to a high-resolution hydrodynamical model, which integrates the actual salinity fluxes.

The box model thus obtained is easily implemented into an existing 0D code, as long as Q_1^{in}, Q_3^{in}, Q_r, V, C_r, C_1, C_2 and C_3 are provided by a 3D model or data.

The high-resolution 3D model diagnostics in the case of the Black Sea Danube plume [4] for these quantities indicate, however, that the box model's precision is limited by several factors:

- A is different when the calculation is based on salinity or heat fluxes, with differences larger than 100%.
- Values of C_2, C_1 and C_3 are highly subjective, especially when assuming, for example, that C_3 is a mixing of C_2 and C. This would amount to a change in the water inflow of Q_3.
- Only volume-averaged temperatures are available, which may be too approximate for use in functions limiting growth rate that are used in biological models. These may react to high temperatures in the surface but not to lower temperatures below the thermocline.

Unfortunately, the mixing coefficient A and the possible mixing of C_2 and C for inflow values C_1 and C_3 may be different for different state variables. This situation may not be overcome, and ultimately implies sensitivity studies in the 0D model, in which these coefficients should be systematically changed and the responses compared.

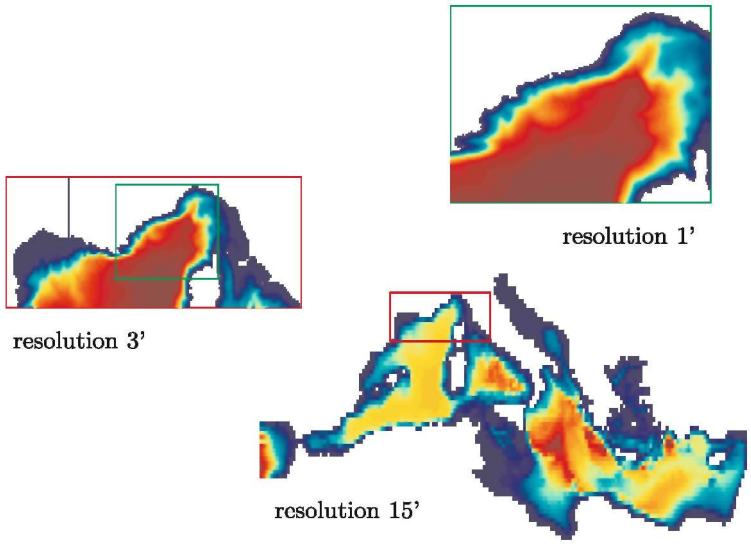
In any case, the fact that integration was performed over a spatial domain clearly involved the need for additional assumptions leading to errors in the box model which, in the case presented here, were quantified [4] using a 3D model.

Generally this is not common (since once a 3D model is available, using a box model instead would disregard too much information) and one should be aware of the inherent limitations of integrated models. Since 3D models are also somehow integrated (at least over the spatial grid boxes and time steps), even such complex models still rely on assumptions related to the unresolved processes (the parameterisation problem). One way to get around this problem in 3D models is to increase resolution in places of interest (Fig. 8), reducing simultaneously the range of scales to be parameterised [6].

5.3 Sediment Processes

Another boundary of particular importance for biochemical components is the bottom boundary, where the sediment layer is acting on the exchanges with the water column. Generally, specific sediment models can be developed and coupled to ocean models. The coupling generally involves flux calculations based on concentrations in the water column (calculated by the ocean model), concentrations in the sediments (given or calculated by a sediment model) and exchange coefficients depending on bottom stress, grain sizes, flocculation effects and other factors influencing re-suspension and deposition of sediments. Here again, intermittency can be highly unpredictable, though strongly modifying exchanges. Among the intermittent processes, sediment avalanches in canyons can significantly enhance shelf-sea exchanges compared to the normal situation in which currents flow along the shelf break.

Fig. 8: *Example of a nested model in the Ligurian Sea*



5.4 Atmospheric Inputs

As for the bottom boundary, the sea surface is another interface with the surrounding systems. Here, intermittency is less of a problem in the sense that high-frequency data are available from atmospheric forecast models. On the other hand, spatial resolution of the surface forcings can be a concern as well as the error bars on imposed heat fluxes. Spatial resolution of atmospheric wind stress strongly controls simulations of upwelling events and subsequent biological model responses. (For example, the mistral and tramontane in the northwestern Mediterranean are controlled by topographic effects hardly represented in global atmospheric models. The corresponding wind stress fields over the Gulf of Lions therefore suffer from decreased sharpness.) Also, errors in heat fluxes might lead to errors in temperature fields or mixed-layer depth and correspondingly changed reactions in biochemical models.

5.5 Estuarine Transformations

Probably the most demanding problem in terms of mathematical modelling is the transition between the river discharge and its arrival in the marine environment. Indeed, in estuaries and low-salinity plumes, biochemical processes are extremely diversified and components undergo a series of biochemical and physical transformations, including flocculation and sedimentation. Therefore, data from river concentrations can rarely be applied directly as an input to the marine model, but must be interfaced by a specialised estuary model.

6. Data Availability

As we have already seen, open boundaries demand the use of information originating from a series of sources, including other models or observations. Another use for observational data is the initialisation of the model, i.e. its spinning up from a given situation. Improper initialisation will generally lead to unreliable results, if not unstable or unphysical solutions, and special care is needed for systems with a long memory of initial conditions.

Unfortunately, obtaining a synoptic high-resolution 3D view of the ocean state (including the tracers) is impossible due to the relative sparseness of in situ data and the cost of generating them. Therefore, data available for model initialisation include a mixture of:

- In situ cruises
- Historical data
- Coastal data
- Satellite data
- Model forecasts of other systems

To use these sources of data for initialisation, the sparse data must be gridded onto the numerical grid of the

model, a step which is called analysis. This is not simply a matter of standard interpolation, but also a matter of filtering out signals in the observations that are not modelled (i.e. in a general circulation model, signals of gravity waves should not be retained during the analysis).

An efficient method to grid data and filter out unresolved scales (noise compared to the scales of interest) is one based on spline approximations, which aims at finding (analysing) a field φ so that it is close to observations but sufficiently regular. It is obtained by minimising a weighting of data-misfit $[d_i - \varphi(x_i, y_i)]^2$ and variability of the analysed field

$$J[\varphi] = \sum_{i=1}^{Nd} \mu_i [d_i - \varphi(x_i, y_i)]^2 + \|\varphi - \varphi_b\|^2 \qquad (14)$$

where the norm $\|\ \ \|$ measures the variability of the field φ on the domain D.

$$\|\varphi\| = \int_D (\alpha_2 \nabla\nabla\varphi : \nabla\nabla\varphi + \alpha_1 \nabla\varphi \cdot \nabla\varphi + \alpha_0 \varphi^2)\ \mathrm{d}D \qquad (15)$$

α_i and μ are parameters of the method and can be obtained to optimally filter out unresolved scales through a calibration of the data-weight μ:

$$\mu = \frac{\varepsilon^2}{\sigma^2} \frac{4\pi}{L^2} \qquad (16)$$

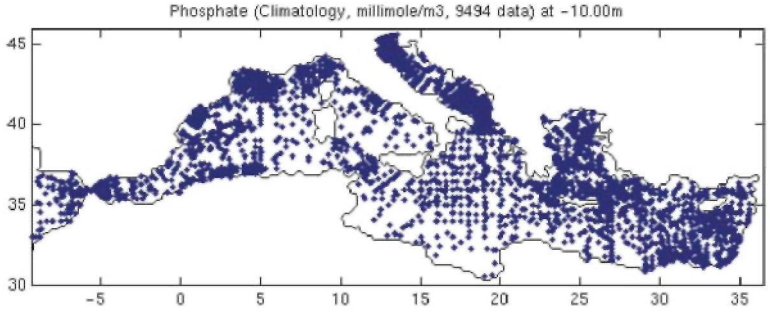
This is generally done by estimating the correlation length L of the data and their signal/noise ratio ε^2/σ^2, where noise includes all signals from all processes not to be dealt with.

Examples of applications of this analysis method are given hereafter. Based on phosphate profiles of the MEDAR database (which also contains historical data on temperature, salinity, chlorophyll, nitrate, silicate, alkalinity etc. [7], http://modb.oce.ulg.ac.be/MEDAR), with unevenly distributed profiles (Fig. 9) over the whole Mediterranean, one can reconstruct [8] an average phosphate concentration (Fig. 10).

Then such data [9] can be used to initialise a biochemical model coupled to a hydrodynamical model which simulates the evolution of the phosphate field (Fig. 11). The evolution during a bloom (Fig. 12) then allows assessment of the upwelled phosphate compared to the riverine input. In the case of the western Mediterranean [9], estimates of the ratio of upwelled phosphates compared to Rhone riverine input were around a factor of 10.

The data analysis methods used to reconstruct fields from irregularly distributed data could also be used to analyse satellite images in which parts of the scenes are covered by clouds. Since data in satellite images are regularly distributed and time-series available, better-suited methods exist which are based on statistical estimates of covariance functions found in the data. Those covariance functions (called empirical since data-based) are then used to analyse the fields under clouds [10]. An example (Fig. 13) of sea surface temperature (SST) filling and validation with MEDAR in situ data showed that errors obtained by the method are comparable to the errors on SST estimations in places without clouds.

Fig. 9: *Data distribution in the MEDAR biogeochemical database for phosphate*



Published in: The Mediterranean Sea (2005), pp. 361-385
Status: Postprint (Author's version)

Fig. 10: *Analysis of phosphate data*

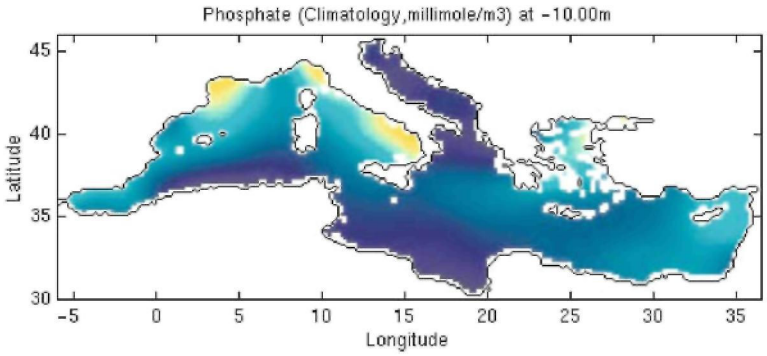


Fig. 11: *Modelled phosphate field before bloom*

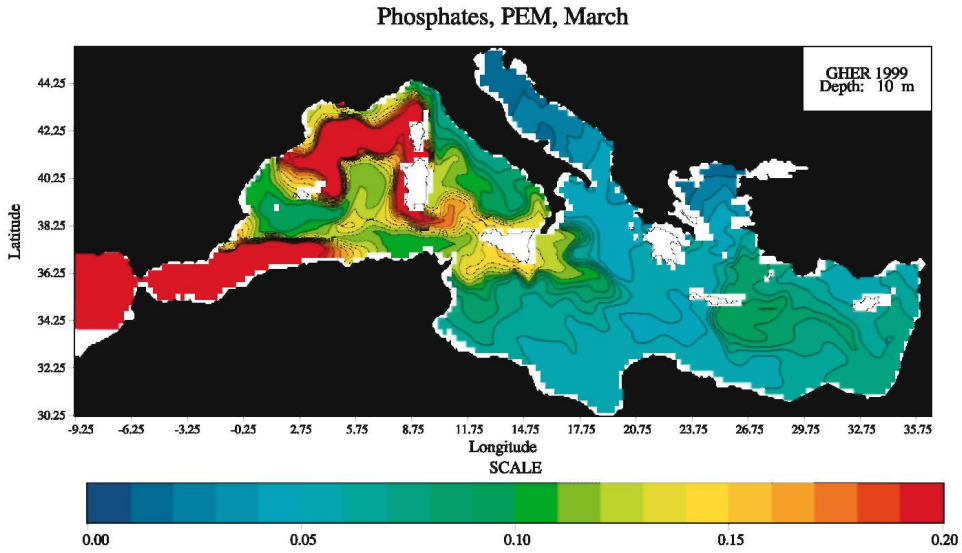


Fig. 12: *Modelled phosphate field after bloom*

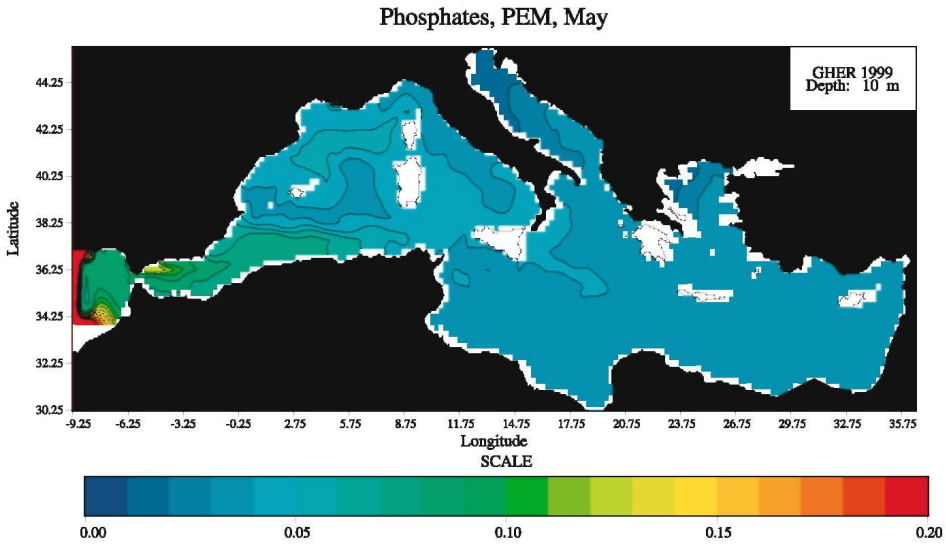
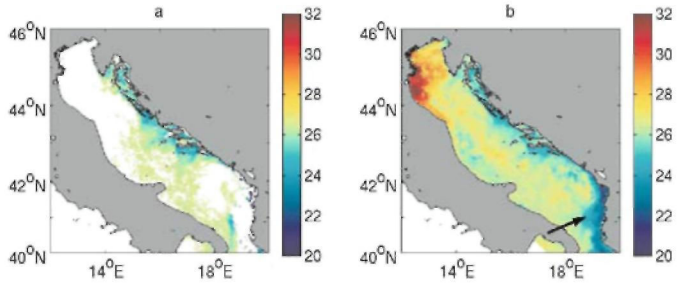


Fig. 13: *Cloud filling of SST data in the Adriatic Sea through empirical orthogonal function (EOF) analysis*



7. Data Assimilation

Data are not only used to initialise and force the model at its boundary, they can also be used to correct model forecasts once new data are available. Basically, when the model provides a forecast of the state vector x^f and observations y^o are available, one would like to combine both in order to get the best estimate of the state of the system x^a.

If we imagine a model with a single forecasted variable T^f and its observation T^o, the best (in statistical terms) estimate of the real value is

$$T^a = \frac{\sigma_o^2}{\sigma_m^2 + \sigma_o^2} T^f + \frac{\sigma_m^2}{\sigma_m^2 + \sigma_o^2} T^o = T^f + \frac{\sigma_m^2}{\sigma_m^2 + \sigma_o^2} \left(T_o - T_f\right) \qquad (17)$$

where σ_m is the standard deviation of the model error and σ_o is the standard deviation of the observational error. This shows simply that we combine two sources of information weighted by the error estimates of the information. The estimated error σ_a of the combination of the two sources of information is then smaller than either model or observational error:

$$\sigma_a^{-2} = \sigma_f^{-2} + \sigma_o^{-2} \qquad (18)$$

The generalisation of this combination of modelled fields and observed fields leads to the so-called Kalman filter, which allows the calculation of the best estimate of the model state x^a:

$$x^a = x^f + K\left(y^o - Hx^f\right) \qquad (19)$$

as a function of the forecasted field x^f, a set of observations y^o not necessarily coinciding with the model state variables, the observation operator H linking the model state vector to observed variables, and the Kalman gain matrix K. The Kalman gain matrix can be calculated if the covariance of the model's error P^f is known and the error covariance matrix of observation R is known (the errors themselves are of course not known, otherwise we would know the exact field; only their statistical distribution is presumed to be known):

$$K = P^f H^T \left(HP^f H^T + R\right)^{-1} \qquad (20)$$

If the error covariance matrices are assumed to be known and model errors are uncorrelated to observation errors, the Kalman filter leads to the best estimate of the state vector and the associated error covariance P^a of the optimal analysis is reduced compared to the error covariance of the model forecast alone

$$P^a = P^f - KHP^f \qquad (21)$$

The error covariance of the model evolves as a function of the model simulation itself (an error in a tracer field is of course advected with the field), and should be updated for each time step according to the dynamics. To do so,

the so-called Lyapounov equation should be solved, which is impossible for most ocean models without further simplifications due to computational limitations. Among the simplifications, using constant model error covariances with prescribed correlation length are the most common. Here [6], we show results of an assimilation of SST in a high-resolution model of the Ligurian Sea (Fig. 14), which allows correction of circulation patterns by assimilating SST (Fig. 15).

Fig. 14: *Effect of data assimilation in the nested Ligurian Sea model. Sea surface temperature (in degrees) (free: model without assimilation; forecast: model forecast one week after last assimilation; analysis: combination of model forecast and observations)*

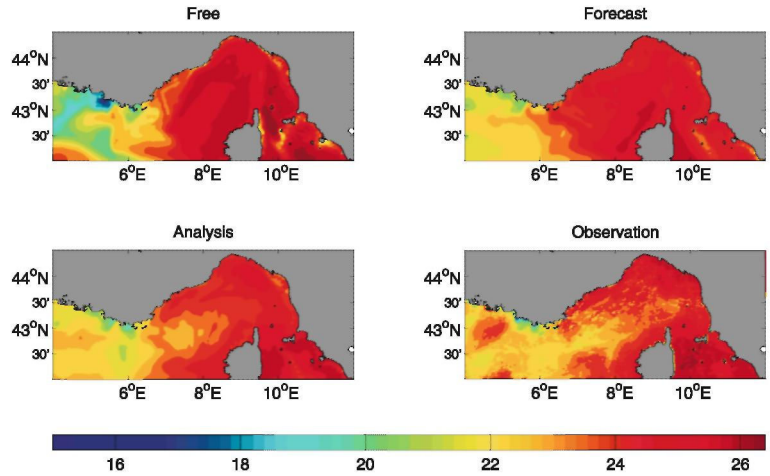
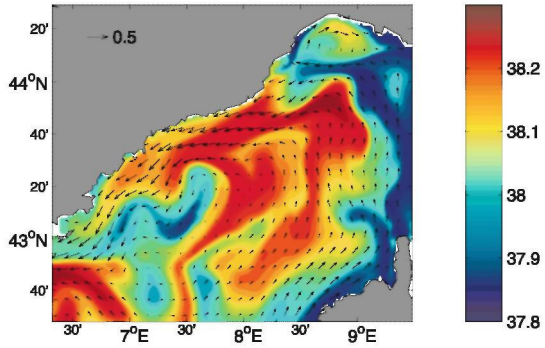


Fig. 15: *Fine resolution model including velocity field showing the northern current and associated surface salinity field*



8. Error Estimations and Appropriate Modelling Approach

The major problem when correcting the model evolution by data assimilation is to know:

• How to quantify the statistical model errors and observational errors (since this information is needed in the analysis step).
• Once specified, if systematic errors are detected during assimilation, how to find the origin of the errors (and subsequently improve the mathematical model).

When characterising errors in the model, we have to observe that they can be of several types:

• Errors in the mathematical model: this error is the difference between the real system evolution and the mathematical solution of the mathematical problem. Examples include errors in interaction laws, parameter values, initial conditions and boundary conditions, topographic or topological simplifications.
• Discretisation errors: this error is the difference between the exact solution of the mathematical continuous problem and the exact numerical solution of the discretised equations.
• Iteration errors: they are the difference between the exact solution of the discrete equations and the numerical

solution obtained by an iterative method (e.g. the convergence error of a Jacobi method solution to a linear algebra problem).
• Rounding errors: errors due to a finite number of digits representing numbers in CPUs.

A well-constructed model should ensure that

rounding errors << iteration errors << discretisation errors
<< modelling errors.

The required inequalities are easily understood. If discretisation errors were larger than modelling errors, there would be no point in telling the numerical model it is an approximation to the physical system we are pretending to describe. If the convergence error was larger than the discretisation error, there would be no point in claiming that the algorithm leads to numerical solutions that satisfy the discretised equations etc.

For assimilation purposes, one can generally neglect convergence errors and rounding errors and include them in general discretisation errors. Therefore only discretisation and model errors should guide the data assimilation.

Error estimates related to discretisations can be performed (for example on flux estimates which are calculated including error bars due to the discretisation [11]), but such an approach is seldom done, and error estimates are now rather based on sensitivity analysis or Monte Carlo approaches. Such a method is used in data assimilation in ensemble approaches, where a series of model simulations (slightly perturbed versions of a reference simulation) are used to derive statistics on model deviations and error covariances. These estimations of P^f are then used in the assimilation process.

9. Validation and Diagnostic Tools

One of the recurrent tasks in modelling is the validation of model results and the pertinent diagnosis of model outputs (in view of the huge amount of data produced by models, appropriate diagnostic tools are a very necessary ingredient of models). Standard statistical tools are of course applicable (rms, correlations, PCA, trends analysis etc.), but sometimes additional model equations can help understanding of the model behaviour. Among the latter possibilities (in addition to adding passive tracers to depict flow patterns), the age theory [12] provides a nice framework to analyse ages of tracers within the model domain, where the age is set to zero at a given location (the inflow for example). The theory needs the calculation of the evolution of a tracer:

$$\frac{\partial c^a}{\partial t} + \nabla \cdot (c^a \boldsymbol{v}) = Q^a - \nabla \cdot (c^a \boldsymbol{m}^a) + \nabla \cdot (\lambda^a \nabla c^a) \qquad (22)$$

and the calculation of the so-called age concentration *a:*

$$\frac{\partial \alpha}{\partial t} + \nabla \cdot (\alpha \boldsymbol{v}) = c^a + q^a - \nabla \cdot (\alpha \boldsymbol{m}^a) + \nabla \cdot (\lambda^a \nabla \alpha) \qquad (23)$$

from which the age *a* of the tracer can be calculated:

$$a = \frac{\alpha}{C} \qquad (24)$$

This quantity gives a direct interpretation of the time a tracer has spent in the domain, subjected to mixing with water parcels of different ages, and is particularly interesting to assess cycling timescales [13].

An illustration is given in the case of the Gulf of Cadiz, in which a nested model was implemented [14,15] to quantify fluxes that are difficult to measure. In particular, the upwelling in the Gulf of Cadiz and the dilution of Tinto-Odiel river discharges before entering the Mediterranean were a concern. For that study, three river discharges (Tinto-Odiel, Guadalquivir, Guadiana) with unit concentrations of tracers were applied. (As a first approximation, nutrients during winter can be described as a non-reacting component described by linear equations. Then concentrations can be scaled by the river input concentration.) The river water discharges used were:

• C_1: Guadiana 157m^3/s
• C_2. Guadalquivir 200 m^3/s

• C_3: Tinto-Odiel 15 m^3/s

The nested model was incorporated in a model covering the Gulf of Cadiz and Alboran Sea (Fig. 16). The circulation revealed the advection and dispersion of the river discharges towards the Gibraltar Strait (Fig. 17) and a sediments tracer (Fig. 18) showed the upwelling along the shelf.

River dilution is easily observed (Fig. 19), but more interestingly, the age calculations (Fig. 20) exhibit the very different timescales of the plume displacements. In particular, the age within the plume can be used as an aid to interpreting biochemical observations within the plume, as it allows quantification of the time already passed since the river water was injected into the marine system. Also, recirculations across the open boundary (resolved by the bidirectional nesting) show the different plume structures and water ages associated with each river.

10. Summary

The modelling of a marine environment is based on a series of assumptions and data requirements that limit the applicability of model results. Care should be taken in designing models appropriate to the question at hand and in ensuring that appropriate data for calibration and validation are available. If parameterisations are used, they should be systematically tested in terms of sensitivity of the modelled solution to changes in values of those parameters. Should the solution change significantly it simply means the parameterised process should not be parameterised but explicitly modelled, for example with nested models.

To perform such sensitivity analyses, estimates on ranges of parameter values as well as ranges of model results are needed, asking for a modeller's insight into the problem. With confidence in the model after validation and sensitivity analysis, the modeller then has at his disposal a very powerful tool that can be exploited systematically in various ways, including statistical analysis of complex model results, standard forecasts, scenario testing and flux quantifications.

Fig. 16: *Circulation modelled in the Gulf of Cadiz at 10 m depth*

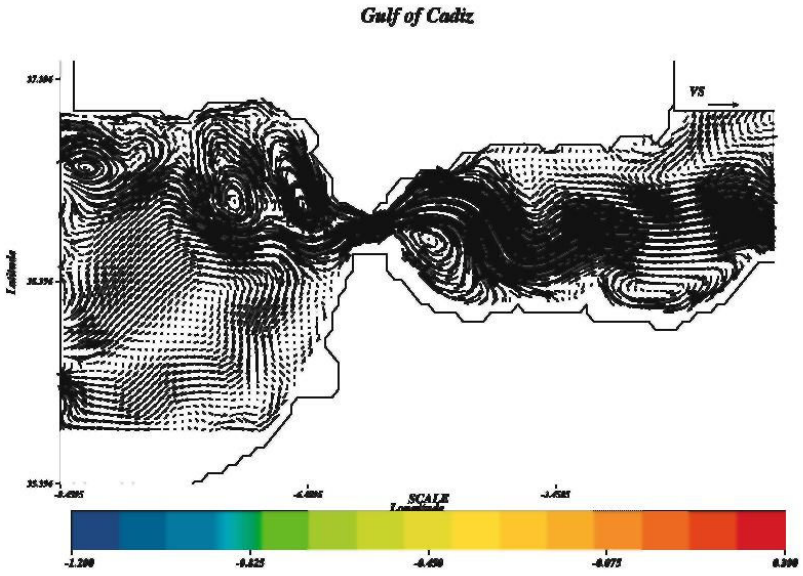


Fig. 17: *Circulation modelled in the Gulf of Cadiz with the nested model at 20 m depth*

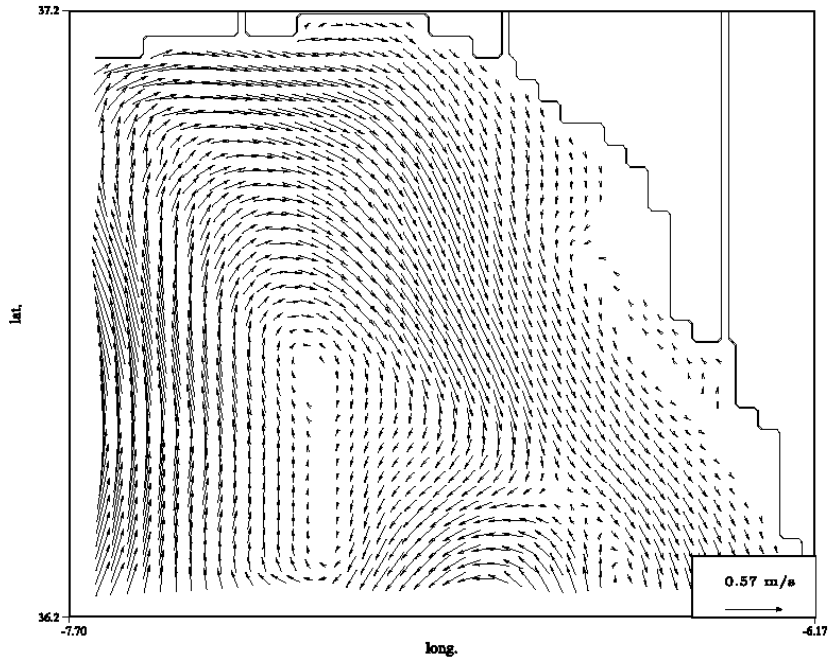
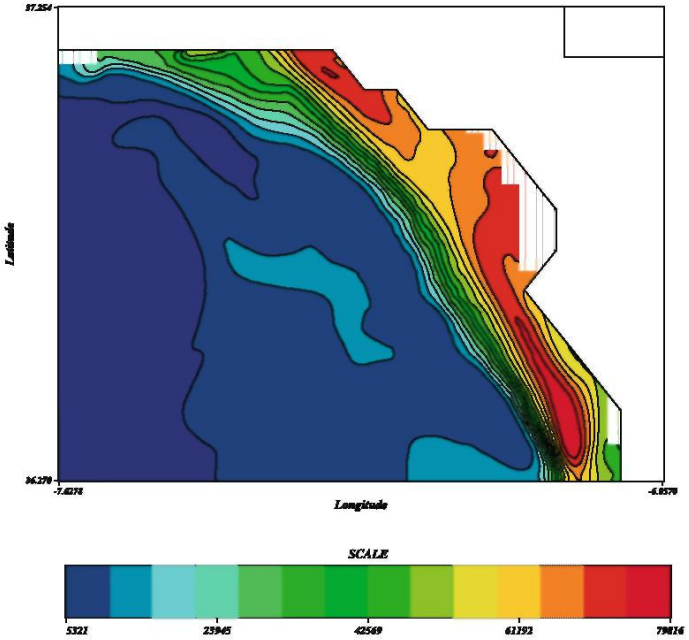


Fig. 18: *Sediment tracer showing upwelled waters at the surface through concentration of a tracer released only on the bottom (arbitrary units)*



Published in: The Mediterranean Sea (2005), pp. 361-385
Status: Postprint (Author's version)

Fig. 19: *Relative concentrations (compared to the input values in ‰) of non-reactive tracers originating from the three rivers Guadiana, Guadalquivir and Tinto-Odiel (from left to right)*

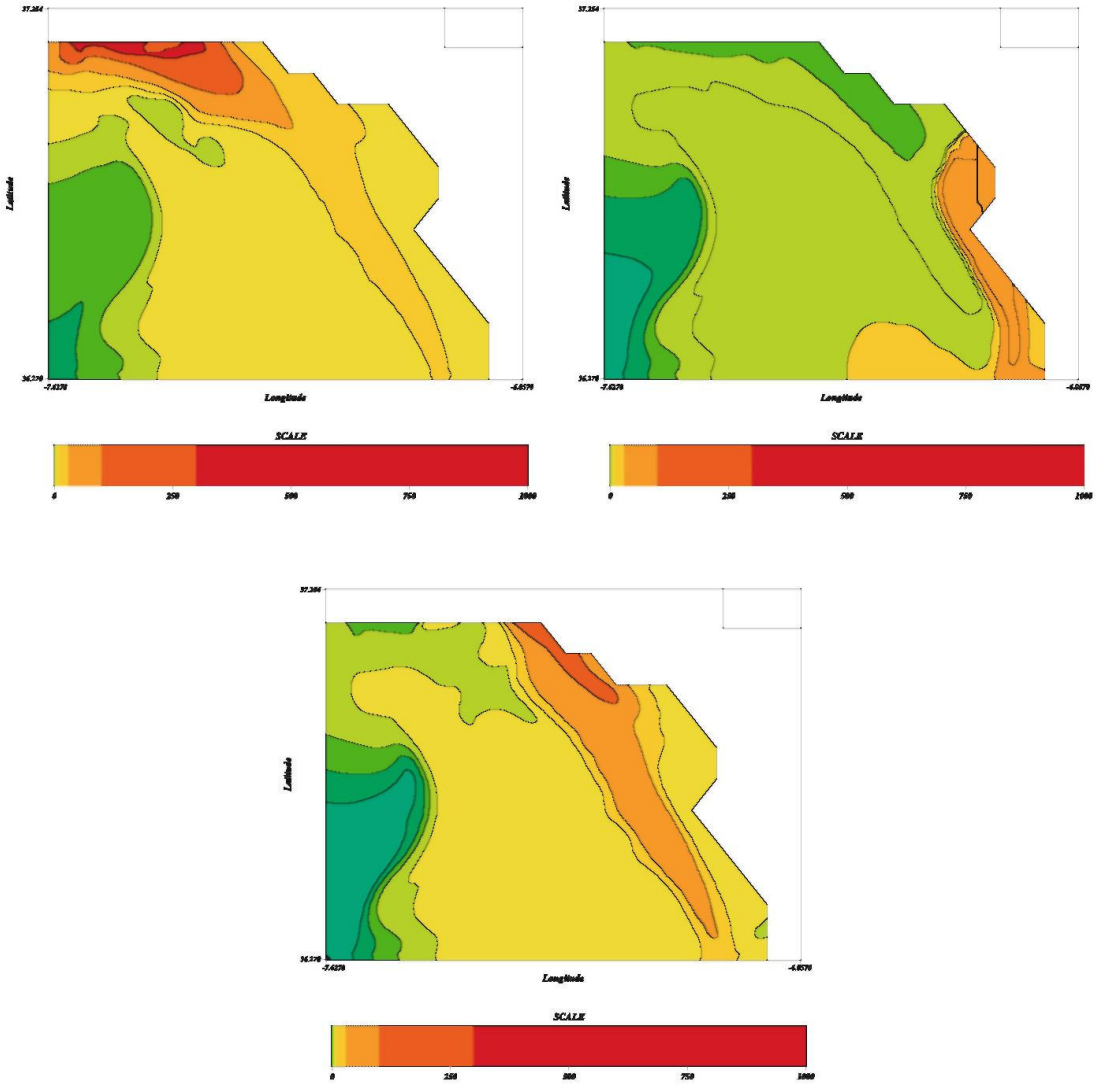
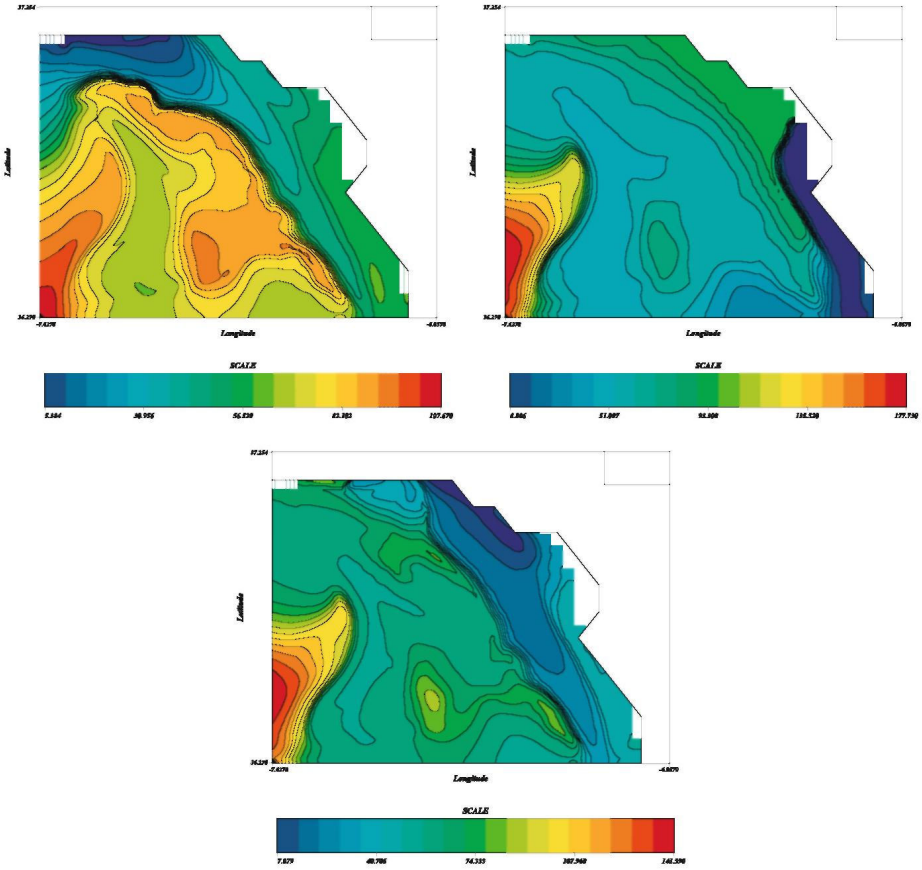


Fig. 20: *River discharge ages. Scales range from 5 (blue) to 110, 180 and 140 days, respectively (red)*



References

1. von Storch H, Zwiers F (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge
2. Beckers J-M (1991) J Marine Syst 1:315
3. Burchard H, Beckers J-M (2004) Ocean Modelling 6:51
4. Beckers J-M, Grégoire M, Nihoul J, Stanev E, Staneva J, Lancelot C (2002) Estuar Coast Shelf Sci 54:453
5. Beckers J-M, Brasseur P, Nihoul J (1997) Deep Sea Res 44:531
6. Barth A, Alvera-Azcarate A, Rixen M, Beckers J-M (2003) Two-way nested model of mesoscale circulation features in the Ligurian Sea. Prog Oceanogr (in press)
7. MEDAR Group (2002) Mediterranean and Black Sea database of temperature, salinity and biochemical parameters and climatological atlas. 4 CD
8. Rixen M, Beckers J-M, Brankart J-M, Brasseur P (2000) Ocean Modelling 2:45
9. Karafistan A, Martin J-M, Rixen M, Beckers J-M (2002) Deep Sea Res 49:67
10. Alvera-Azcárate A, Barth A, Rixen M, Beckers J (2005) Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions. Application to the Adriatic Sea. Ocean Modelling 9:325
11. Grégoire M, Beckers J-M (2004) Biogeosciences 1:33
12. Deleersnijder E, Mouchet A, Delhez E, Beckers J-M (2002) Math Comput Model 36:121
13. Delhez E, Lacroix G, Deleersnijder E (2003) Ocean Dynamics
14. Elbaz-Poulichet F, Braungardt C, Achterberg E, Morley N, Cossa D, Beckers J-M, Nomerange P, Cruzado A, Leblanc M (2001) Cont Shelf Res 21:1961
15. Elbaz-Poulichet F, Morley N, Beckers J-M, Nomerange P (2001) Marine Chem 73:193