
Assimilation de données dans un modèle d'écosystème marin de la mer Ligure Data assimilation in a marine ecosystem model of the Ligurian Sea

Stéphanie Magri^a, Pierre Brasseur^{b*} and Geneviève Lacroix^c

^a Ifremer, BP 70, 29280 Plouzané, France

^b LEGI/CNRS, BP 53X, 38041 Grenoble, France

^c MUMM, 100 Gulledelle, 1200 Bruxelles, Belgique

*pierre.brasseur@hmg.inpg.fr

Abstract: The objective is to explore the potentialities of sequential statistical estimation methods to assimilate observations in a primary production biological model coupled to a vertical 1D hydrodynamical model characterised by a k-l turbulent closure. The assimilation method is derived from the SEEK filter (Singular Evolutive Extended Kalman filter), which uses an error subspace represented by multivariate empirical orthogonal functions (EOFs). Real data assimilation experiments collected at sea have been realised to reconstruct the variability of the Ligurian Sea ecosystem during the FRONTAL field experiment. To cite this article: S. Magri et al., C. R. Geoscience 337 (2005).

Keywords: Physical/biogeochemical model; Numeric modelling; Data assimilation; Kalman filter; Ligurian Sea

Data assimilation in a marine ecosystem model of the Ligurian Sea.

Data assimilation have been conducted in a one-dimensional, coupled physical-ecosystem model of the upper ocean in order to improve the realism of the simulations and to predict the biogeochemical state of the ocean in a realistic way.

The coupled model (Modèle d'ECOsystème du G.H.E.R. et du L.O.V. – MODECOGeL) simulates the primary production in a coastal zone of the Ligurian Sea in the North Western Mediterranean Sea, where oligotrophic conditions prevail. The ecosystem dynamics is represented by twelve compartments expressed in nitrogen units. The model takes into account nitrate and ammonium to distinguish new production from regenerated production, three compartments of phytoplankton (pico-, nano- and microzooplankton) and three compartments of zooplankton (nano-, micro- and mesozooplankton), two types of particulate organic nitrogen, dissolved organic nitrogen and bacteria for a representation of the microbial loop. The coupling with an hydrodynamic model determines the physical constraints associated to the development of a seasonal mixed layer. The stratification of the water column and the associated vertical turbulent diffusivities, are key parameters of the evolution of the marine ecosystem. The model is constrained by heat fluxes and wind stress at the air-sea interface according to real meteorological data. The coupled system has been developed and validated on the basis of field data collected during the *FRONTAL* campaigns between 1984 and 1988 at two stations located along a transect between Nice and Calvi.

The assimilation method is based on the Singular Evolutive Extended Kalman (SEEK) filter, which is a reduced-order assimilation scheme. The error sub-space is decomposed into multivariate Empirical Orthogonal Functions (EOFs) of the system's variability. The SEEK filter computes sequential corrections of the model trajectories, taking into account the

balance between the confidence of the model prediction and the accuracy of the observed quantities. The experiment presented here concerns nitrate and chlorophyll-a profiles assimilation. In the usual case, the error subspace EOFs are evaluated using the model variability as a proxy of the system variability. But in this work the originality is to use EOFs deduced from observations. This approach permits to introduce processes which could not be represented by a vertical 1D model alone.

1. Introduction

Le besoin de représenter le milieu marin de façon la plus complète et la plus réaliste possible a conduit la communauté scientifique à développer un grand nombre de modèles numériques, donnant accès à la compréhension des processus dynamiques et biologiques, et à des informations complémentaires aux observations de terrain (données satellitales, *in situ*...). Dans cet esprit, le modèle MODECOGeL (*MOD*èle d'*ECO*ystème du *G.H.E.R.* et du *L.O.V*) [1] a été développé en vue de représenter l'écosystème marin de la mer Ligure, située en Méditerranée Occidentale. Il s'agit d'un modèle unidimensionnel vertical de production primaire couplé à un modèle de couche de mélange océanique. Ce dernier a fait l'objet d'une validation détaillée [1, 2] à partir de données prélevées le long d'une radiale reliant Nice à Calvi au cours des différentes missions FRONTAL [3] réalisées entre le 12 mars 1984 et le 19 décembre 1988. Des conditions de vent et de flux de chaleur, évaluées sur base de données météorologiques réelles, sont imposées au modèle hydrodynamique. Le modèle biologique est quant à lui basé sur le cycle de l'azote, décrit par douze variables d'état incluant plusieurs classes de taille planctoniques et une description explicite de la boucle microbienne. Le couplage entre ces deux types de modèles est nécessaire afin de prendre en compte l'influence de la dynamique sur la répartition verticale des variables biologiques. A noter cependant que le couplage dont il est question ici n'est réalisé que dans un seul sens, dans la mesure où aucun mécanisme de rétroaction de la biologie sur l'hydrodynamique n'est pris en compte. Bien que le modèle MODECOGeL simule de façon satisfaisante l'écosystème de la mer Ligure, ce dernier présente un certain nombre de limitations, résultant de l'hypothèse d'homogénéité horizontale, de l'usage de forçages atmosphériques imparfaits ou encore de paramétrisations approximatives des processus dynamiques. Ainsi, Lacroix ([1]) a pu identifier une tendance généralisée au réchauffement ainsi qu'une dérive en sel sur 5 années consécutives de simulation, une sous-estimation des concentrations en nitrate ou encore des

maxima chlorophylliens trop élevées. Lacroix et Grégoire ([2]) mettent également en évidence une surestimation de la profondeur de la nitracline dans les simulations, probablement due à l'absence de processus d'advection verticale par le modèle 1D.

Une piste pour corriger ces tendances consiste à appliquer, parallèlement au modèle, une méthode d'assimilation de données qui a pour objectif de fournir une représentation plus réaliste du système en utilisant de manière optimale toutes les informations provenant du modèle et des observations. L'assimilation de données est réalisée au moyen du filtre SEEK (Singular Evolutive Extended Kalman filter) [4], méthode dérivée du filtre de Kalman qui utilise un principe de réduction d'ordre afin de rendre l'algorithme praticable sur le plan numérique. Dans nos expériences, la matrice de covariance des erreurs du premier estimé P_0 est calculée à partir d'une décomposition en Fonctions Empiriques Orthogonales (EOFs) multivariées issues de la variabilité des observations utilisées pour l'assimilation [5]. Contrairement à une approche plus classique qui consiste à prendre en compte la variabilité du modèle pour une représentation des EOFs, cette approche originale permet d'introduire des informations complémentaires sur les processus qui ne peuvent pas être représentés par une modélisation unidimensionnelle verticale, et ainsi rendre les simulations plus réalistes.

Plusieurs expériences d'assimilation de données réelles ont donc été réalisées, montrant l'importance d'une estimation cohérente de la matrice P_0 ([5]). L'expérience présentée par la suite concerne l'assimilation conjointe de profils de nitrate et de chlorophylle_a. L'année 1986 a été choisie car il s'agit de la période annuelle la mieux décrite par les données biologiques acquises au cours des campagnes FRONTAL.

2. Le modèle MODECOGeL

L'étude s'appuie sur le modèle MODECOGeL, un modèle d'écosystème marin couplé à un modèle de couche de mélange océanique de la mer Ligure. C'est un modèle unidimensionnel

vertical développé dans le but d'étudier les processus biologiques le long de la colonne d'eau et de quantifier les flux biogéochimiques en fonction des conditions météorologiques et hydrodynamiques du milieu. Il a été calibré et validé sur base des données FRONTAL prélevées entre 1984 et 1988 à deux stations sur les six réparties selon un transect reliant Nice à Calvi, ces deux stations étant situées de part et d'autre du front Liguro-Provençal. Elles correspondent respectivement à la station 1, site des programmes DyFAMeD et JGOFS-France, située au large à 28 milles nautiques de Cap Ferrat, et la station 6, station côtière à 3 milles nautiques de Cap Ferrat. Pour notre part, la zone d'étude correspond à la station 6, les forçages atmosphériques y étant plus réalistes par rapport à la station du large et le jeu de données plus important. En outre, il sera question d'étudier le comportement du modèle qui devra intégrer, via l'assimilation de données, des informations qui ne sont pas directement représentées par la modélisation 1D verticale, comme par exemple la signature des processus horizontaux caractérisant cette station.

2.1. Le modèle hydrodynamique

Le modèle MODECOGeL est un modèle de couche de mélange multi-niveaux, non linéaire, avec une fermeture turbulente de type k-l. Le modèle hydrodynamique prend en compte 5 variables d'état, à savoir, la température, la salinité, les deux composantes horizontales de la vitesse u et v , et l'énergie cinétique turbulente, cette dernière variable permettant la fermeture du schéma de turbulence. Les conditions aux limites à l'interface air-mer du modèle hydrodynamique sont imposées sous forme de flux de chaleur, d'eau douce et de quantité de mouvement estimés à partir de données météorologiques mesurées toutes les 3 heures aux stations du Cap Ferrat et de l'aéroport de Nice [6]. Ces conditions limites déterminent au premier ordre l'activité turbulente dans la couche mélangée, la diffusion turbulente conditionnant très fortement la répartition spatiale des composantes biologiques dans la

colonne d'eau, surtout dans le cadre d'un modèle unidimensionnel vertical. L'importance du pilotage du modèle biologique à partir des informations fournies par le modèle hydrodynamique se situe donc principalement au niveau de la prise en compte des variations spatio-temporelles du coefficient de diffusion turbulente. En outre, la température, en tant que variable pronostique, va permettre d'une part l'apparition d'une thermocline saisonnière qui conditionne fortement l'activité phytoplanctonique de la mer Ligure, et d'autre part de considérer la dépendance en température de nombreux processus biologiques.

Le modèle hydrodynamique est représenté par les équations primitives ci-dessous. Elles sont ramenées à la dimension verticale en supposant l'homogénéité horizontale [6].

$$\begin{aligned}
 \frac{\partial u}{\partial t} &= \frac{\partial}{\partial z} \left(\tilde{\nu} \frac{\partial u}{\partial z} \right) + fv \\
 \frac{\partial v}{\partial t} &= \frac{\partial}{\partial z} \left(\tilde{\nu} \frac{\partial v}{\partial z} \right) - fu \\
 \frac{\partial T}{\partial t} &= \frac{\partial}{\partial z} \left(\tilde{\lambda}^T \frac{\partial T}{\partial z} \right) \\
 \frac{\partial S}{\partial t} &= \frac{\partial}{\partial z} \left(\tilde{\lambda}^S \frac{\partial S}{\partial z} \right) \\
 \frac{\partial k}{\partial t} &= \tilde{\nu} \left\| \frac{\partial \mathbf{V}}{\partial z} \right\|^2 (1 - Rf) - \varepsilon + \frac{\partial}{\partial z} \left(\tilde{\nu} \frac{\partial k}{\partial z} \right)
 \end{aligned}$$

où $\mathbf{V} = (u, v)$ est la vitesse horizontale, f la fréquence de Coriolis, T la température, S la salinité, k l'énergie cinétique turbulente, $\tilde{\nu}$ et $\tilde{\lambda}^y$ les coefficients de viscosité et de diffusion turbulente, ε le taux de dissipation de l'énergie cinétique turbulente et Rf le nombre de Richardson de flux mesurant l'importance relative de la production et de la destruction de l'énergie cinétique turbulente.

2.2. Le modèle biologique

Comme le modèle biologique a été développé dans le but d'étudier le rôle de la dynamique sur la répartition verticale des nutriments et de la production primaire, seuls les premiers

niveaux trophiques de la chaîne alimentaire sont modélisés. Il prend ainsi en compte 12 variables d'état, à savoir le nitrate et l'ammonium afin de distinguer la production nouvelle de la production régénérée, 3 classes de taille phytoplanctoniques (pico-, nano- et microphytoplancton), 3 classes de taille zooplanctoniques (nano-, micro- et mésozooplancton), la matière organique particulaire de taille 1 et 2, l'azote organique dissous et les bactéries. C'est un modèle relativement simple, inspiré du modèle de Fasham et al ([7]) avec en plus la prise en compte d'une boucle microbienne, afin d'être représentatif de la zone d'étude. La figure 1 illustre les processus biologiques régissant les flux de matière entre les différents compartiments du modèle.

Comme nous l'avons signalé auparavant, la répartition spatiale des différents organismes biologiques sera conditionnée par le calcul d'un coefficient de diffusion turbulente. Ainsi, le couplage s'effectue principalement par l'intermédiaire d'un terme de diffusion turbulente $\tilde{\lambda}$ dans les équations biologiques [1] :

$$\frac{\partial \mathbf{Bio}}{\partial t} = \frac{\partial}{\partial z} \left(\tilde{\lambda} \frac{\partial \mathbf{Bio}}{\partial z} \right) + \mathbf{sources} - \mathbf{pertes}$$

où $\tilde{\lambda} = \mathbf{f}(\tilde{\mathbf{v}})$

En ce qui concerne les conditions aux limites, le flux en surface de toutes les variables biologiques est nul. A la base de la colonne d'eau modélisée, un flux nul est imposé à toutes les variables du modèle couplé (hydrodynamique et biologique). Les équations aux dérivées partielles sont résolues par une méthode aux différences finies sur une grille décalée. Le pas de temps est de 10 minutes et la discrétisation verticale est de 1 mètre entre la surface et une profondeur de 200 mètres.

3. L'assimilation de données

Le problème de base que nous cherchons à résoudre par l'assimilation de données est une meilleure estimation spatio-temporelle de certaines propriétés de l'océan, en combinant de

façon optimale les solutions d'un modèle numérique et des observations. En effet, un modèle est toujours affecté d'un certain nombre de limitations et les observations sont souvent très éparpillées dans l'espace et le temps. En utilisant conjointement les solutions du modèle et des observations, nous visons à une représentation aussi réaliste que possible de l'océan.

3.1 Méthodes d'assimilation en biologie marine

Deux principales approches sont possibles afin de résoudre le problème d'estimation optimale par l'assimilation de données : l'approche variationnelle et l'approche séquentielle [8]. Celle utilisée dans le cadre de ce travail est de type séquentiel, dérivée de la théorie de l'estimation statistique optimale. Elle consiste à fournir un estimé de l'état du système en corrigeant séquentiellement la prévision du modèle chaque fois que des observations sont disponibles. L'approche statistique séquentielle fait appel à une connaissance des erreurs correspondant au modèle et aux observations afin de minimiser l'incertitude sur la solution estimée au sens des moindres carrés.

Kalman ([9]) a proposé un schéma statistique séquentiel optimal afin de résoudre les problèmes inverses linéaires. La particularité du filtre de Kalman est de propager dans le temps une estimation statistique des covariances d'erreurs pour des systèmes dynamiques linéaires. Evensen ([10]) a quant à lui développé un filtre de Kalman d'Ensemble s'appliquant à des modèles dynamiques non linéaires comportant de très nombreux degrés de liberté. Une autre approche, mise en place par Pham et al ([4]) concerne le filtre SEEK, méthode utilisée dans le cadre de ce travail et plus longuement développée par la suite. Très peu d'études concernant l'utilisation d'une approche séquentielle avec des modèles couplés physiques-biogéochimiques ont été menées jusqu'à présent. Nous pouvons cependant citer les travaux de Eknes et Evensen ([11]) qui utilisent un filtre de Kalman d'Ensemble dans un modèle unidimensionnel à trois composantes biologiques, ou le projet européen DIADEM

(Development of advanced data assimilation systems for operational monitoring and forecasting of the North Atlantic and Nordic Seas) où le filtre SEEK et le filtre de Kalman d'Ensemble ont été implémentés dans un modèle tridimensionnel couplé, l'objectif étant de contrôler l'ensemble des variables biologiques au moyen d'observations SeaWiFS de la couleur de l'eau ([12, 13]).

L'approche variationnelle, que nous n'utiliserons pas dans le cadre de ce travail, est dérivée de la théorie du contrôle optimal [14]. Le principe repose sur la minimisation d'une fonction coût, dépendant d'un certain nombre de paramètres qui contrôlent la trajectoire du modèle. Cette fonction coût mesure les écarts quadratiques entre un jeu de données et une solution du modèle. Les paramètres de contrôle peuvent correspondre aux paramétrisations qui caractérisent les processus du modèle, ou à des conditions limites ou initiales par exemple. Cette approche est celle qui est le plus souvent utilisée dans les modèles couplés physiques-biogéochimiques [15, 16, 17, 18, 19, 20, 21]. En effet, la détermination des paramètres qui interviennent dans la paramétrisation des processus biologiques est un problème récurrent en modélisation. Cependant, l'assimilation variationnelle présente un certain nombre de limitations [5], dont une majeure : elle est beaucoup plus difficile à mettre en œuvre pour contrôler l'état du système couplé en agissant simultanément sur toutes les causes identifiables des erreurs qui affectent les composantes hydrodynamiques et biologiques du modèle. D'où l'intérêt d'une approche séquentielle, à partir de laquelle il est en pratique plus facile d'agir sur l'ensemble des variables d'état du modèle dynamique et du modèle biologique.

3.2 Le filtre SEEK

Le filtre SEEK [4] est un filtre de Kalman simplifié, dans lequel l'espace d'état complet est projeté dans un espace d'état réduit. La matrice de covariance des erreurs au temps initial P_0 est représentée par un nombre limité de Fonctions Empiriques Orthogonales (EOFs) multivariées décrivant les modes dominants de la variabilité du système, et définissant de cette façon la structure de l'espace réduit au temps initial. L'approche multivariée est utile afin de propager l'information depuis les quantités observées vers les quantités non observées. En effet, les observations ne correspondent qu'à un nombre limité de variables du modèle. A travers la représentation multivariée des covariances d'erreurs, les corrections affectent non seulement les variables observées, mais également l'ensemble des variables du modèle. Cette approche est renforcée par le résultat d'expériences qui ont mis en avant le degré de sous-optimalité du schéma d'analyse consécutif au calcul de corrections monovariées [5].

Le filtre SEEK repose sur un algorithme qui enchaîne une étape d'analyse et une étape de prévision. L'étape d'analyse corrige l'état du système en intégrant les observations disponibles à l'instant où la correction est effectuée, dans le but de réduire les écarts entre les observations et les prédictions du modèle, en fonction de leurs erreurs respectives. Une covariance des erreurs d'analyse est diagnostiquée. On effectue ensuite une étape de prévision qui consiste en l'intégration de l'état du système par le modèle à partir des corrections effectuées au préalable. En parallèle, la covariance des erreurs de prévision est calculée par intégration de la covariance des erreurs de l'analyse précédente via la dynamique du modèle. Successivement, sont appliquées une étape d'analyse puis une étape de prévision, d'où le caractère récurrent de la méthode.

4. Assimilation des données biologiques FRONTAL

Le principal objectif des expériences d'assimilation est d'obtenir une représentation la plus complète possible de la variabilité de l'écosystème marin durant un cycle annuel, en combinant les informations issues des données FRONTAL avec l'utilisation du modèle MODECOGeL. L'assimilation de données peut également être interprétée comme une extrapolation ou une interpolation dynamique des observations selon la dimension temporelle. L'intégration de données par assimilation dans la simulation couplée permet en principe de corriger les défauts du modèle qui résultent de l'hypothèse d'homogénéité horizontale, de l'usage de forçages atmosphériques imparfaits ou encore de paramétrisations approximatives des processus dynamiques. La validation du modèle couplé avait en effet fait ressortir quatre principales limitations, touchant aussi bien la partie hydrodynamique que la partie biologique du modèle. Il s'agit d'une tendance généralisée au réchauffement et d'une dérive en sel au cours du temps, d'une sous-estimation des concentrations en nitrate et de maxima chlorophylliens trop élevés [1]. Nous allons explorer ici la question de l'assimilation conjointe de profils verticaux de nitrate et de chlorophylle_a, le modèle hydrodynamique n'étant pas contrôlé par des observations de nature physique (température et salinité).

4.1. Les difficultés rencontrées

Les données FRONTAL utilisées ont été échantillonnées à certaines profondeurs : 0, 5, 10, 20, 30, 50, 75, 100, 150 et 200 mètres. Pour mener à bien l'assimilation, la répartition spatiale des mesures doit en principe tenir compte de la variabilité verticale des quantités observées. Or des expériences préliminaires ont montré que l'échantillonnage spatial choisi lors des missions FRONTAL n'était pas optimum par rapport à la variable 'nitrate', et dans une moindre mesure à la chlorophylle_a [5]. En outre, la répartition des données dans le temps doit aussi pouvoir décrire la variabilité temporelle avec suffisamment de précision. Or nous ne

disposons d'aucun prélèvement de nitrate entre le 08 juillet et le 02 septembre 1986 et de chlorophylle_a entre le 18 juin et le 02 septembre 1986, bien que cette année ait été choisie car elle comptabilise le plus de profils sur les 5 années de campagne. Ce déficit au niveau de l'échantillonnage temporel rend a priori l'assimilation plus délicate.

Une difficulté supplémentaire apparaît lors de l'assimilation de données de chlorophylle_a. En effet, cette quantité représente une agrégation entre plusieurs compartiments du modèle, à savoir les trois catégories phytoplanctoniques, et il nous est impossible de connaître directement les proportions relatives de chacune de ces catégories à partir d'une simple donnée de chlorophylle_a. Cette difficulté est résolue dans l'algorithme SEEK grâce à une représentation de l'opérateur d'observation \mathbf{H} qui permet de relier les données de chlorophylle_a aux concentrations des 3 variables phytoplanctoniques, au prorata de leur répartition statistique moyenne estimée à partir de la variabilité naturelle du modèle libre. La variabilité du rapport Chl/Phy est donc prise en compte au travers de l'importance relative des 3 catégories phytoplanctoniques représentées dans le modèle.

Notons enfin que les données dont nous disposons ne représentent qu'une partie des variables d'état biologiques. Il nous est donc impossible de valider le modèle biologique dans son ensemble, et notamment le zooplancton et la matière organique, sachant que nous ne disposons pas de données relatives à ces variables. Nous nous appuyons donc sur des représentations spatio-temporelles par interpolation des données FRONTAL, ainsi que des écarts RMS (au sens des moindres carrés) entre d'une part les observations et une simulation sans assimilation, et d'autre part les observations et la simulation avec assimilation.

4.2. Protocole des expériences d'assimilation

La version du filtre SEEK utilisée pour les expériences d'assimilation est dite « fixe », à savoir que la statistique d'erreur n'évolue pas d'une étape d'analyse à la suivante selon la

dynamique du modèle, mais est simplement amplifiée au moyen d'un facteur constant fixé à 2. Cette amplification de l'erreur au cours de l'étape de prévision est une paramétrisation qui globalise à la fois les effets de croissance dynamique de l'erreur d'analyse résiduelle et l'incrément d'erreur de simulation lié aux imperfections du modèle ([4]). La durée d'un cycle d'assimilation varie en fonction de la distribution temporelle des observations qui est d'une quinzaine de jours en moyenne. L'erreur d'observation est supposée non corrélée spatialement et son écart-type est fixé à de $0,3 \text{ mmolN/m}^3$ pour le nitrate, et à $0,2 \text{ mmolN/m}^3$ pour la chlorophylle (après conversion des unités). A noter cependant que ces valeurs ont été fixées par ajustement de façon à tenir compte à la fois de l'erreur de mesure et de l'erreur de représentativité des données assimilées (qui dépend notamment de la « richesse » de la base réduite utilisée dans le filtre SEEK). Seules les variables biologiques définissant le vecteur d'estimation sont corrigées statistiquement à chaque cycle, vu la difficulté de spécifier des covariances d'erreur suffisamment robustes entre toutes les variables du modèle couplé. Le vecteur d'estimation considéré ici prend en compte les deux catégories de nutriment et les trois classes phytoplanctoniques.

Dans les simulations examinées ici, les conditions initiales en température, salinité, azote et phytoplancton sont imposées au 1er janvier 1986 sous forme d'un profil vertical moyen calculé à partir des données FRONTAL représentatives de la période hivernale. Les autres variables d'état sont initialisées à une valeur constante sur la verticale, selon la procédure décrite dans [2].

4.3. Assimilation de données de nitrate et de chlorophylle_a

Rappelons que le but de cette expérience est de pouvoir corriger des concentrations en nitrate trop faibles et des maxima chlorophylliens trop élevés, en intégrant des informations sur les processus non représentés par le modèle MODECOGeL via les modes d'erreurs et les

observations. L'originalité de l'approche développée repose sur l'utilisation de modes d'erreurs déduits des observations [5] et non d'une simulation de référence comme cela est classiquement effectué [12]. Les EOFs représentent la variabilité du nitrate et des 3 catégories phytoplanctoniques.

La figure 2 montre une représentation d'une moyenne annuelle ainsi que des EOFs calculées d'une part à partir de la variabilité du modèle et d'autre part à partir de la variabilité des données FRONTAL de l'année 1986, entre 0 et 200 mètres de profondeur. Les sorties du modèle MODECOGeL ont été prises en fonction de l'échantillonnage spatio-temporel des données FRONTAL. Les 3 premiers modes de variabilité représentant 93,03 % de la variance correspondant à la variabilité issue des simulations du modèle libre, et 97,15 % de la variance représentée dans les données FRONTAL. La valeur du pourcentage indiqué correspond à la somme des valeurs propres des EOFs retenues divisée par la somme de toutes les valeurs propres des EOFs issues du jeu de données de départ.

Au regard des profils présentant une moyenne annuelle, on note une augmentation des concentrations en nitrate avec la profondeur, avec cependant une sous-estimation des valeurs simulées. Quant à la chlorophylle_a, le modèle simule un pic de concentration vers 80 mètres de profondeur, non visible d'après les données FRONTAL. D'autre part, en dessous de 120 mètres de profondeur, les simulations tendent à sous-estimer les concentrations. Les modes calculés via le modèle montrent une faible variation en nitrate entre la surface et 80 mètres de profondeur, contrairement aux second et troisième modes résultant des données FRONTAL. Le premier mode présente des profils similaires, excepté en dessous de 160 mètres de profondeur. Pour ce qui est de la chlorophylle_a, la première EOF calculée via les données FRONTAL montre une certaine homogénéité le long de la colonne d'eau, contrairement à l'EOF calculée à partir du modèle. Cependant, d'après l'étude de la fonction temporelle associée au premier mode de variabilité [5], la variabilité saisonnière est relativement bien

représentée. Par contre, les second et troisième modes ont des signaux plus complexes à interpréter, et on ne trouve pas de concordance prononcée entre les deux types de mode, ce qui confirme une certaine divergence entre les données réelles et les résultats du modèle.

La figure 3 montre une représentation des écarts RMS entre les simulations assimilées (étape de prévision = 'xf' ; étape d'analyse = 'xa') et les données FRONTAL d'une part, et les simulations sans assimilation (libre) et les données FRONTAL d'autre part, selon la profondeur et le temps. Les expériences sont initialisées au 1^{er} janvier 1986. Les corrections sur le nitrate sont très marquées, et les écarts RMS en dessous de 100 mètres de profondeur sont fortement atténués grâce à l'assimilation. En effet, à 200 mètres de profondeur, les écarts RMS entre la simulation sans assimilation et les observations sont de l'ordre de 2,5 mmolN/m³, alors qu'ils ne sont plus que de 0,5 mmolN/m³ lors de l'analyse. Au cours du temps, ces mêmes écarts ne dépassent pas 0,5 mmolN/m³. Le niveau d'erreur associé à l'étape de prévision se situe entre l'analyse et la simulation libre, ce qui indique que les corrections effectuées durant l'étape d'analyse ne sont pas totalement dégradées lors de l'étape de prévision, et que l'échantillonnage temporel des données ne semble pas être si insuffisant. Les corrections sur la chlorophylle_a sont moins marquées, bien que significatives, notamment entre la surface et 80 mètres de profondeur. Au cours du temps, on note la présence d'un certain nombre d'anomalies (matérialisés par des pics) générées par la simulation libre ; ces dernières sont atténuées grâce à l'assimilation.

D'après les données FRONTAL (figure 4), on assiste en février à des remontées en nitrate dont les concentrations oscillent entre 0,50 et 0,75 mmolN/m³. Le modèle libre simule également ces remontées, mais les concentrations n'atteignent pas celles prédites par les observations. De plus, entre 150 et 200 mètres de profondeur, les valeurs en nitrate sont sous-estimées. La simulation assimilée tend à corriger ces anomalies, et on peut noter une évolution spatio-temporelle plus conforme à la réalité, avec des concentrations corrigées entre

150 et 200 mètres de profondeur, même si les concentrations en début d'année sont encore sous-évaluées. Quant à la chlorophylle_a (figure 5), les données FRONTAL indiquent une floraison entre 0 et 50 mètres de profondeur en mars et avril, dont le maximum de concentration atteint $0,70 \text{ mgChl/m}^3$. A contrario, malgré des quantités de nitrate moins importantes début février dans la simulation libre, on assiste à une poussée de plus grande envergure par rapport aux observations. Les maxima chlorophylliens dans ce cas de figure peuvent alors atteindre $1,13 \text{ mgChl/m}^3$. L'assimilation atténue ce phénomène et permet une meilleure représentation spatio-temporelle de la chlorophylle_a, notamment à compter du mois de septembre.

5. Conclusion

Des expériences d'assimilation de données ont été réalisées au moyen d'un modèle couplé biologique-hydrodynamique unidimensionnel vertical de la mer Ligure et d'un filtre SEEK, qui correspond à une méthode d'estimation statistique dérivée du filtre de Kalman. L'expérience présentée concerne l'assimilation de profils verticaux de nitrate et de chlorophylle_a prélevés au cours de l'année 1986. L'objectif principal de cette expérience est de représenter de la façon la plus réaliste possible les simulations du modèle lors d'un cycle annuel, le modèle couplé présentant un certain nombre d'imperfections.

Au vu des résultats de l'expérience décrite, nous pouvons affirmer que l'assimilation conjointe de profils biologiques de nitrate et de chlorophylle_a conduit à des améliorations significatives. Les corrections se sont avérées assez sensibles à la manière dont est construite la matrice de covariance des erreurs initiale. L'approche présentée ici fut d'initialiser le sous-espace d'erreurs au moyen d'EOFs représentatives de la variabilité décrite par des observations réelles, ce qui permet d'obtenir des résultats plus satisfaisants qu'en utilisant des EOFs calculés à partir de la variabilité naturelle du modèle ([1]). Cette approche à l'avantage

d'intégrer via les différents modes d'erreurs des processus non représentés par la modélisation unidimensionnelle verticale. Ceci suppose des stratégies d'observations adéquates, qui donnent accès à l'information multivariée relative à cette variabilité. Dans notre cas, si de futurs prélèvements étaient à effectuer, il serait intéressant d'effectuer d'avantage de prélèvements en dessous de 100 mètres de profondeur afin d'obtenir de plus amples informations sur la variabilité en nitrate, notamment durant les remontées de nutriments depuis les couches profondes vers la couche superficielle au cours des trois premiers mois de l'année, ainsi qu'au printemps et en automne dans les 100 premiers mètres à plus haute fréquence temporelle durant les floraisons phytoplanctoniques.

De futurs travaux pourraient s'appuyer sur l'assimilation conjointe de données satellitales haute fréquence de la couleur de l'eau à des données *in situ*, afin d'extrapoler une information assimilée en surface le long de la colonne d'eau. L'assimilation de profils de température et de salinité a également été mise en œuvre, en vue de corriger les imperfections du modèle hydrodynamique et ainsi essayer de rendre les simulations du modèle biologique plus réalistes [5]. L'assimilation conjointe de données de nature physique et biologique n'a pas encore été exploitée, et il serait très intéressant d'effectuer des travaux dans cette voie afin de corriger l'ensemble du modèle couplé.

Littérature citée

- [1] G. Lacroix, Simulation de l'écosystème pélagique de la mer Ligure à l'aide d'un modèle unidimensionnel. Etude du bilan de matière et de la variabilité saisonnière, interannuelle et spatiale, thèse, université de Liège, 1998.
- [2] G. Lacroix, M. Grégoire, Revisited ecosystem model (MODECOGeL) of the Ligurian Sea : seasonal and interannual variability due to atmospheric forcing, *J. Mar. Syst.* 37 (4) (2002) 229-258.
- [3] P. Nival, Résultats des missions 'FRONTAL'. Tome I : 1984-1985. Tome II : 1986-1988, FRONTAL report, 1989.
- [4] D.T. Pham, J. Verron, M.C. Roubaud, A singular evolutive extended Kalman filter for data assimilation in oceanography, *J. Mar. Syst.* 16 (1998) 323-340.
- [5] S. Magri, Assimilation de données dans un modèle d'écosystème marin couplé à un modèle de couche de mélange océanique de la mer Ligure, thèse, université Joseph Fourier, 2002.
- [6] G. Lacroix, S. Djenidi, Extending the G.H.E.R. 3D Model to the modelling of Ecosystems in Western Mediterranean Coastal Zones : Results from an Exploratory Study, Proceedings of the third EROS 2000 Workshop on Research in the North Western Mediterranean Sea, Commission of the European Communities, Texel (1992).
- [7] M.J.R. Fasham, H.W. Ducklow, S.M. McKelvie, A nitrogen-based model of plankton dynamics in the oceanic mixed layer, *J. Mar. Res.* 48 (1990) 591-639.
- [8] F. Boutier, P. Courtier, Data assimilation concepts and methods, ECMWF report, 1999.
- [9] R.E. Kalman, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* 82D (1960) 35-45.
- [10] G. Evensen, Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, *J. Geo. Res.* 99 (C5) (1994) 10143-10162.
- [11] M. Eknes, G. Evensen, A ensemble Kalman filter with a 1-D marine ecosystem model, *J. Mar. Syst.* 36 (2002) 75-100.
- [12] V. Carmillet, J.M. Brankart, P. Brasseur, H. Drange, G. Evensen, J. Verron, A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical-biochemical model of the North Atlantic ocean, *Ocean Modelling* 2 (2001) 45-60.
- [13] L.J. Natvik, G. Evensen, Assimilation of ocean color data into a biochemical model of the North Atlantic. Part I : Data assimilation experiments, *Ocean Modelling* 2 (2001) 19-44.
- [14] J.L. Lions, Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles, Dunod, 1968.

- [15] B. Faugeras, M. Lévy, L. Mémerly, J. Verron, J. Blum, I. Charpentier, Can biogeochemical fluxes be recovered from nitrate and chlorophyll data? A case study assimilating data in the Northwestern Mediterranean Sea at the JGOFS-DYFAMED station, *J. Mar. Syst.* 40-41 (2003) 99-125.
- [16] J. Gunson, A. Oschlies, V. Garçon, Sensitivity of ecosystems parameters to simulate satellite ocean color data using a coupled physical-biological model of the North Atlantic, *J. Mar. Res.* 57 (1999) 613-639.
- [17] Y.H. Spitz, J.R. Moisan, M.R. Abbott, J.G. Richman, Data assimilation and a pelagic ecosystem model : parameterization using time series observations, *J. Mar. Syst.* 16 (1998) 51-68.
- [18] P. Prunet, J.F. Minster, D. Ruiz-Pino, I. Dadou, Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean : 1. Method and preliminary results, *Global Biogeochem. Cycles*, 10 (1) (1996a) 111-138.
- [19] P. Prunet, J.F. Minster, V. Echevin, I. Dadou, Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean : 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate and pCO₂ data, *Global Biogeochem. Cycles*, 10 (1) (1996b) 139-158.
- [20] G.C. Hurtt, R.A. Armstrong, A pelagic ecosystem model calibrated with BATS and OWSI data, *Deep Sea Res. I* 46 (1999) 27-61.
- [21] M. Schartau, A. Oschlies, J. Willebrand, Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method, *Deep Sea Res. II* 48 (2001) 1769-1800.

Légende des figures

Figure 1 : Représentation schématique du modèle biologique MODEDOGeL utilisé dans les simulations couplées ; les 12 compartiments représentent les variables d'état biologiques suivantes: nitrate et ammonium (NO₃, NH₄) ; pico- (Pp), nano- (Np) et micro-phytoplancton (Mp) ; nano- (Nz), micro- (Miz) et meso-zooplancton (Mez) ; matière organique particulaire (POM1, POM2) ; azote organique dissoute (DON) ; bactéries (Bac). Les processus biologiques responsables des flux entre les compartiments sont décrits à gauche.

Figure 1 : Schematic representation of the MODEDOGeL biological model used in the coupled simulations; the 12 compartments represent the biological state variables : nitrate and ammonium (NO₃, NH₄) ; pico- (Pp), nano- (Np) and micro-phytoplankton (Mp) ; nano- (Nz), micro- (Miz) and meso-zooplankton (Mez) ; particulate organic matter (POM1, POM2) ; dissolved organic nitrogen (DON) ; bacteria (Bac). The biological processes responsible for the fluxes between the compartments are described on the left.

Figure 2 : Représentation, en fonction de la profondeur (mètres), d'une moyenne annuelle et des 3 premiers modes de variabilité concernant la variable nitrate et la chlorophylle-a.

Figure 2 : Representation, according to the depth (meters), of an annual average and the first three modes of variability of nitrate and chlorophyll.

Figure 3 : Ecart RMS, en fonction de la profondeur (mètres) et du temps (jours), de la variable nitrate (mmolN/m³) et de la chlorophylle-a (mgChl/m³). Courbe noire : simulation sans assimilation. Courbes grises : prévision (trait continu) et analyse (pointillés).

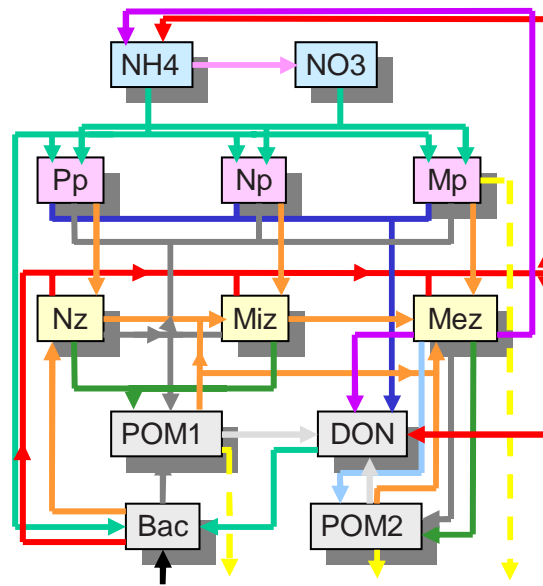
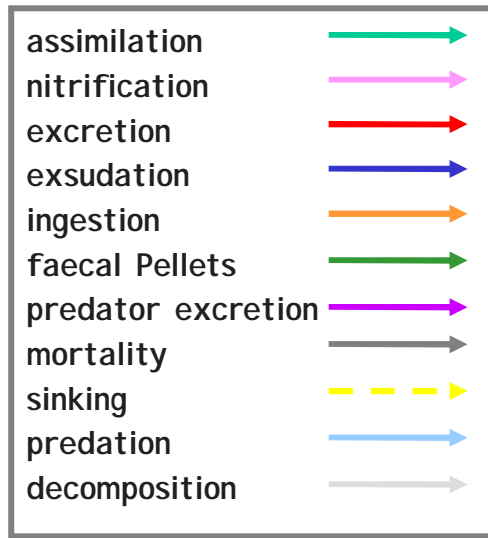
Figure 3 : RMS error, according to the depth (meters) and the time (days), of nitrate (mmolN/m³) and chlorophyll (mgChl/m³). Black curve : simulation without assimilation. Grey curves : forecast (continue line) and analysis (dot line).

Figure 4 : Evolution spatio-temporelle de la variable nitrate (mmolN/m³). Figure du haut : données FRONTAL. Figure du milieu : simulation libre. Figure du bas : simulation assimilée.

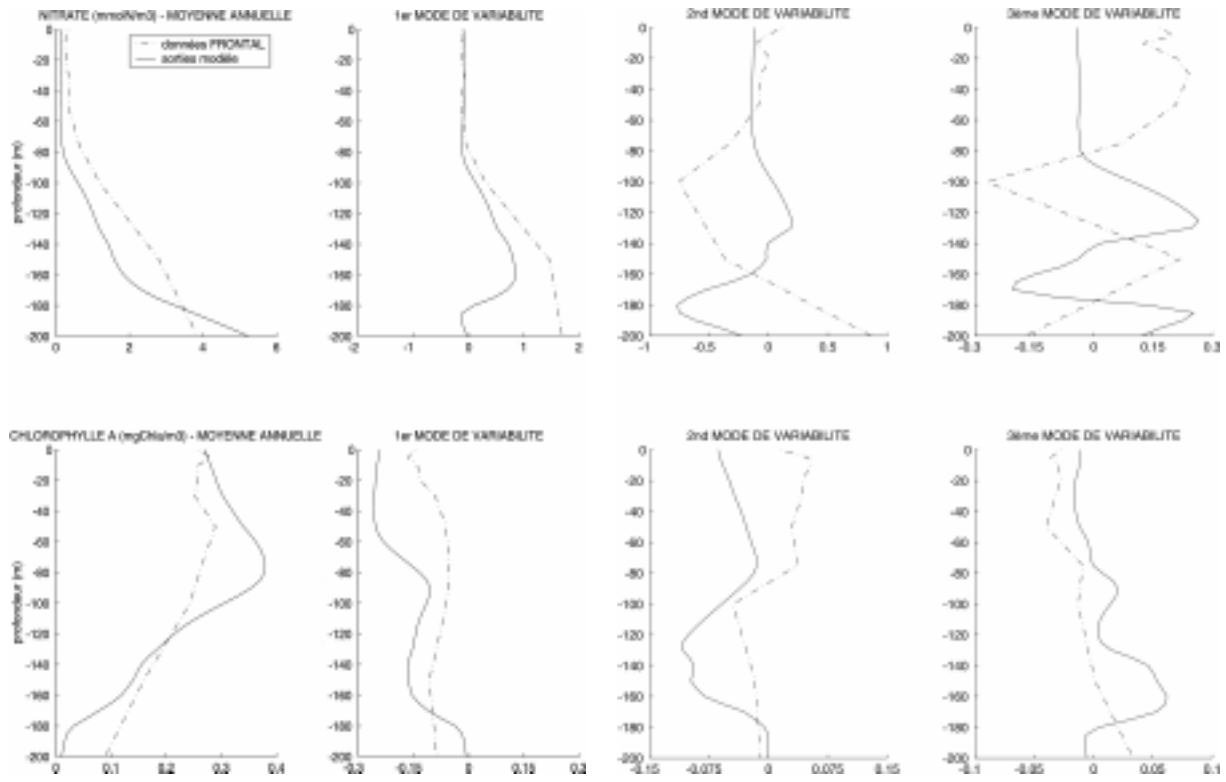
Figure 4 : Spatio-temporal evolution of the nitrate (mmolN/m³). Top panel : FRONTAL data. Middle panel : free model run. Bottom panel : simulation with assimilation.

Figure 5 : Evolution spatio-temporelle de la chlorophylle-a (mgChl/m³). Figure du haut : données FRONTAL. Figure du milieu : simulation libre. Figure du bas : simulation assimilée.

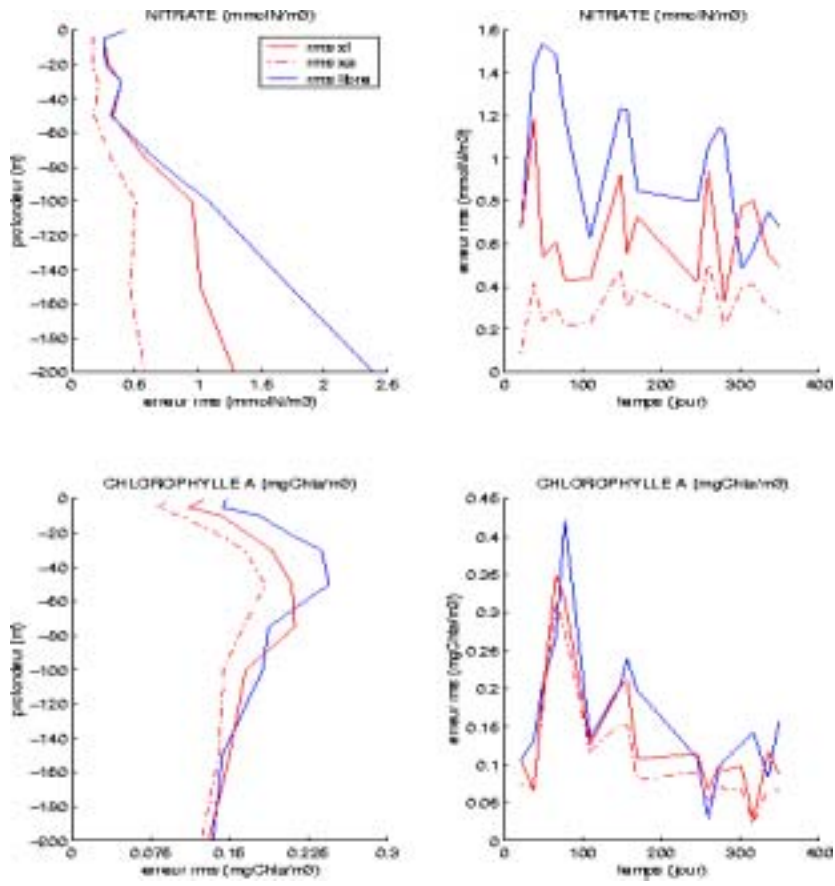
Figure 5 : Spatio-temporal evolution of the chlorophyll (mgChl/m³). Top panel : FRONTAL data. Middle panel : free model run. Bottom panel : simulation with assimilation.



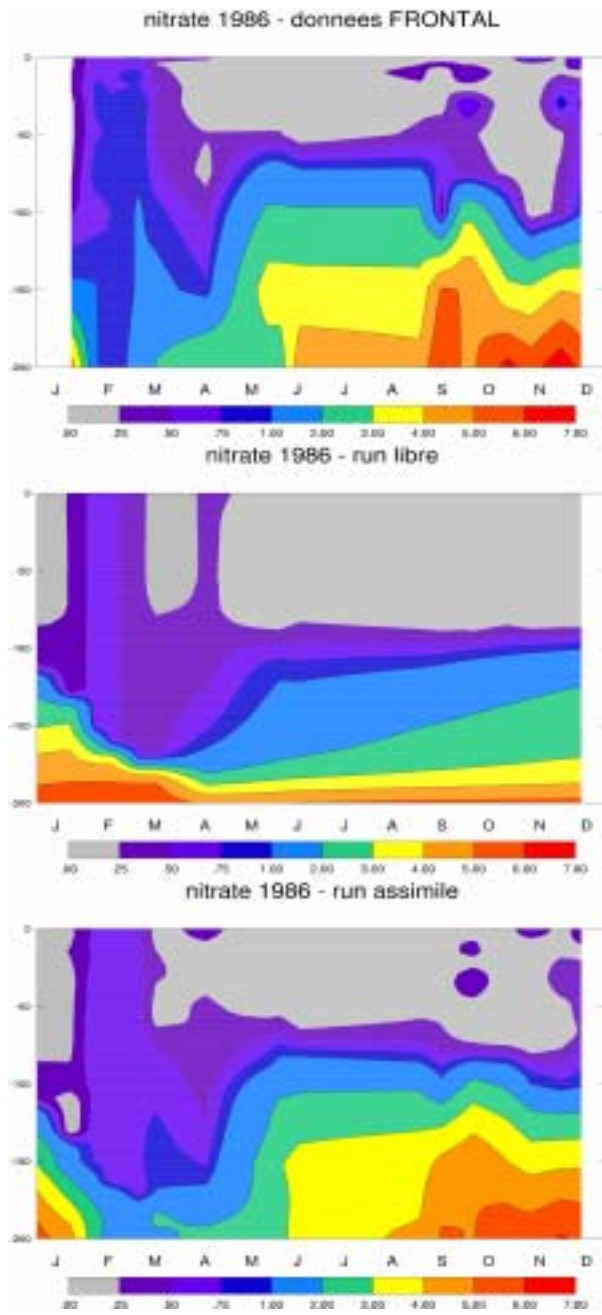
Magri *et al.* – fig. 1



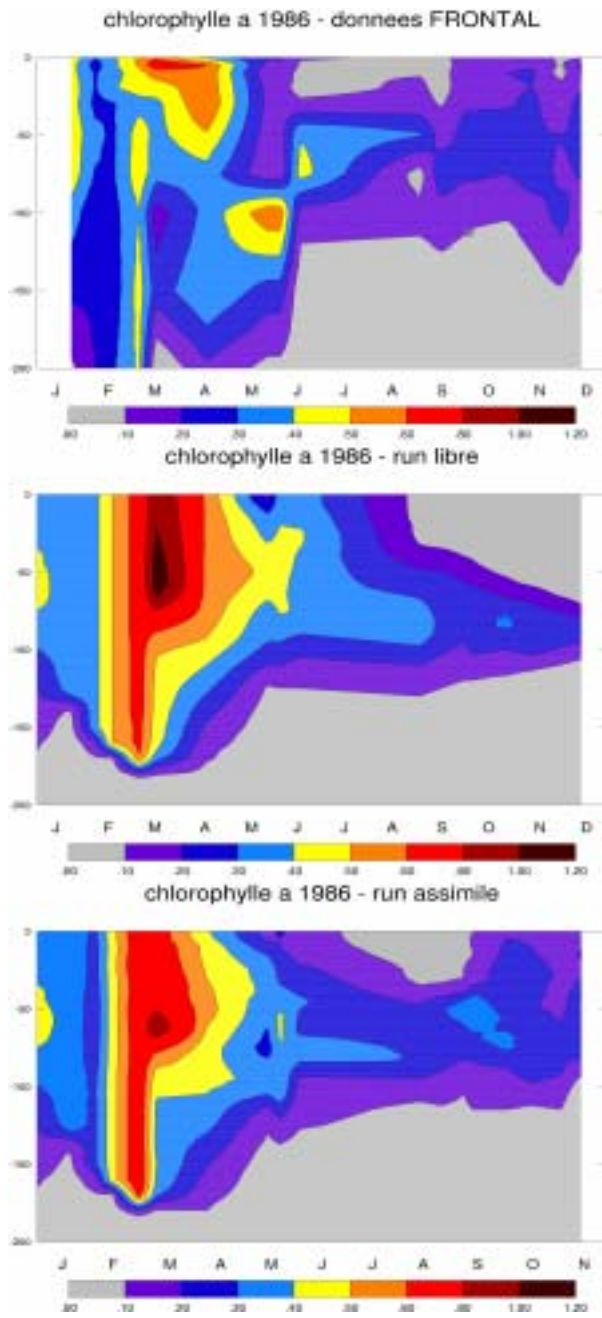
Magri *et al.* – fig. 2



Magri *et al.* – fig. 3



Magri *et al.* – fig. 4



Magri *et al.* – fig. 5