**Western and
Central Pacific
Fisheries
Commission**

**SCIENTIFIC COMMITTEE
FIFTH REGULAR SESSION**

10-21 August 2009
Port Vila, Vanuatu

A USER MANUAL FOR SEAPODYM VERSION 2.0: application with data assimilation (draft version)

WCPFC-SC5-2009/EB-IP-13

Patrick Lehodey[1], and Inna Senina[1]

[1] Marine Ecosystems Modeling and Monitoring by Satellites, CLS, Space Oceanography Division, 8-10 rue Hermès, 31520 Ramonville, France

A User manual for

SEAPODYM Version 2.0

application with data assimilation
(draft version)

P. Lehodey, I. Senina

August 3, 2009

SEAPODYM is a model developed initially for investigating spatial tuna population dynamics under the influence of both fishing and environmental effects. The model is based on advection-diffusion-reaction equations. The main features of this model are i) forcing by environmental data (either observed or modeled), ii) prediction of both temporal and spatial distribution of age-structured populations, iii) prediction of total catch and size frequency of catch by fleet when fishing data (catch and effort) are available, iv) parameter optimization based on fishing data assimilation techniques.

This modelling effort started in 1995 at the Secretariat of the Pacific Community in Noumea, New Caledonia, under two consecutive EU-funded projects: SPR-TRAMP (1995-2000) and PROCFISH (2002-2005). The model development also benefited of a grant from the PFRP (Pelagic Fisheries Research Program) of the University of Hawaii, allowing the implementation of irregular grids and initiating the work for parameter optimization (2004-05). Since 2006, the development has continued within the MEMMS section (Marine Ecosystem Modeling and Monitoring by Satellites) of the Spatial Oceanography Division of CLS, a subsidiary of the French CNES and IFREMER Institutes. Collaboration with SPC continues, with funding support from a new EU-funded SPC project, SCIFISH, as well as with the PFRP through a second grant (2006-09) for the project "'Climate and Fishing Impacts on the Spatial Population Dynamics of Tunas (Project no. 657425)"'.

An enhanced version of the model (SEAPODYM 2.0) is now ready for use. It includes revised mechanisms due to changes in the modelling of mid-trophic organisms of the pelagic ecosystem with several pelagic mid-trophic functional groups. Dynamics of tuna populations have been also revised with expanded definitions of habitat indices, movements, and natural mortality. The code for parameter optimization has been completed and tested.

Following a general introduction and a brief review of the main mechanisms used in the model, this user's manual provides the information needed to run numerical simulations for tuna or tuna-like species, and the method to fit the model predictions to observations by estimating model parameters, investigation of parameter observability, improving initial conditions and the model functionality.

---

[1]Marine Ecosystems Modeling and Monitoring by Satellites, CLS, Satellite Oceanography Division. 8-10 rue Hermes, 31520 Ramonville, France

[2]Oceanic Fisheries Programme, Secretariat of the Pacific Community, BPA5 Noumea, New Caledonia

[3]Pelagic Fisheries Research Program, University of Hawaii at Manoa. 1000 Pope Road, Honolulu, HI 96822 United States

Contents

Introduction

The Spatial Ecosystem and Population Dynamics model (SEAPODYM) model has been continuously enhanced to provide a general framework allowing integration of biological and ecological knowledge of tuna species, and potentially other oceanic top-predator species, within a comprehensive description of the pelagic ecosystem. It has been upgraded to include more detailed relationships between the population dynamics and basic biological and ecological functions, including a realistic representation of the vertical oceanic habitat, both in terms of physical and foraging conditions. SEAPODYM is an age-structured population model describing spatial and temporal dynamics of top predator species in a direct link with forage (prey) biomass and environmental variability. The forage fields are predicted by separate sub-model in which various mid-trophic level organisms are classified through their dial migration pattern and the spatiotemporal transfer of energy from oceanic primary productivity to the preys of tunas is described using allometric scaling equation and passive transport of the biomass with oceanic waters. The environmental variables, which drive tuna dynamics (temperature, currents, oxygen and primary production) are predicted by coupled physical-biogeochemical models. The model also includes a rigorous mathematical parameter estimation procedure using available catch and size frequency data.

Because the model is driven by the bio-physical environment of the ecosystem, it was possible to reduce the number of parameters that describe the complete spatially-explicit population dynamics of a species to twenty-one, i.e., a small number relative to the number of variables described in the model. A few more parameters should be added to include the growth function that is still provided by independent studies in this version. Other parameters concern the description of fisheries (selectivity and catchability).

The model is parameterized through assimilation of commercial fisheries data, and optimization is carried out using maximum likelihood estimation approach. For parameter optimization, we implemented adjoint methodology to obtain an exact, analytical evaluation of the likelihood gradient. The approach to select the "'best parameter estimate"' is based on a series of computer experiments in order i) to determine model sensitivity with respect to variable parameters and, hence, investigate their observability, ii) to estimate observable parameters and their errors, and iii) to justify the reliability of found solution. The new version of the model is fully detailed in recent papers (Lehodey et al., 2008; Senina et al., 2008).

Chapter 1

The model

1.1 Modelling approach

When considering the definition of an ecosystem, i.e., how assemblages of species are organized in space and time, and how they interact with each other and the physical environment, modelling the ocean pelagic ecosystem is obviously a challenge that requires drastic simplifications. These simplifications need to be considered carefully in close relation with the level of observation and knowledge that we have for each component of the system to be sure that the model can be properly developed and parameterized and then evaluated by comparing prediction to available observations. In addition, the model is focusing on the population dynamics and the fisheries of exploited species as there is a special interest to provide a new generation of modeling tools for the management of these species, taking into account not only the impact of the fisheries but also the natural fluctuations of the populations in their climate-driven ecosystem.

While the level of information for the physical and biogeochemical components is now sufficient to make possible detailed three-dimensional realistic planetary scale simulations of ocean physic and lower trophic (phytoplankton) level biology, it is clear that the information for the intermediate trophic levels (micronekton) is a weak link. Organisms in these groups that provide the forage of tuna and other oceanic top predators consist of macrozooplankton and micronekton, typically crustaceans, fish, and cephalopods, with sizes in the range of 2-20 cm.

1.1.1 Mid-Trophic Level Model

The modelling of mid-trophic organisms of the pelagic ecosystem is a critical step in linking the coupled physical-biogeochemical models to population dynamics of large pelagic predators. Modelling this component of the system then requires an approach as simple as possible but that can be linked, however, to the other components without loosing the benefit of the better knowledge and detailed modelling possibility they offer. Many of these mid-trophic

species have optimized their behavior under the constraints of finding food in the productive upper layers while avoiding predators as far as possible, i.e., migrating in the upper layers after the sunset and leaving it during sunrise, so that vertical distributions and migrations are finally structuring all the pelagic foodweb. It seems therefore appropriate to define intermediate trophic levels with several simple large functional groups with a dynamic linked to the one provided by ocean-biogeochemical models.

We have developed a modelling approach with six different groups characterised by their vertical behavior, i.e., occurence of diel migration between epipelagic, mesopelagic and bathypelagic layers (Fig. 1.1). Parameterization of the dynamics of these components is based on a temperature-linked time development relationship. Then, a simple energy transfer from primary production is used, justified by the existence of constant slopes in log-log biomass size spectrum relationships. Recruitment, ageing, mortality and passive transport with horizontal currents, taking into account vertical behavior of organisms, are modeled by a system of advection-diffusion-reaction equations. This part of the model is not detailed in this version of the manual. A full description of the Mid-Trophic Model is provided in Lehodey et al. (*in press*).

1.1.2 Tuna and tuna-like population model

Top predators in the marine pelagic ecosystem are essentially opportunistic omnivorous predators. Their diets reflect both the faunal assemblage of the component of the ecosystem that they explore and their aptitude to capture prey species at different periods of the day (i.e., daytime, nightime, twilight hours). It seems that most of them are in the upper layer during the night. But high sensory specialisation (e.g. olfaction in sharks, vision in bigeye tuna, swordfish and cephalopods or echolocation in marine mammals), and morphological and physiological adaptations (e.g., thermo-regulation) allow them also to exploit the dark and colder deeper layers. The 3-layer ocean definition used for the mid-trophic (forage) species seems to match particularly well with the known vertical behaviour of large predators (Fig. 1.2).

Large pelagic predators are often exploited species for which there are detailed knowledge on biology and physiology, population structure, ecology and fishing data. Thus, it is possible to develop a detailed model of the spatial dynamics of these predator species.

The modeling approach for the tuna population dynamics has been first described in Bertignac et al. (1998), and modified in Lehodey (2001), Lehodey et al. (2003). The enhanced version presented in this document is described in Lehodey et al. (2008) and Senina et al. (2008). Populations are age-structured and movements described with an advection-diffusion equation. Diffusion is used to represent undirected random movement, i.e., kinesis, while in contrast advection allows to describe directed movement in response to some external stimulus, i.e., taxis. Here, the directed movement is following the gradient (slope) of a habitat index (klinotaxis). Therefore, the movement of these top-predator species is modeled as a combination of kinesis and klinotaxis.
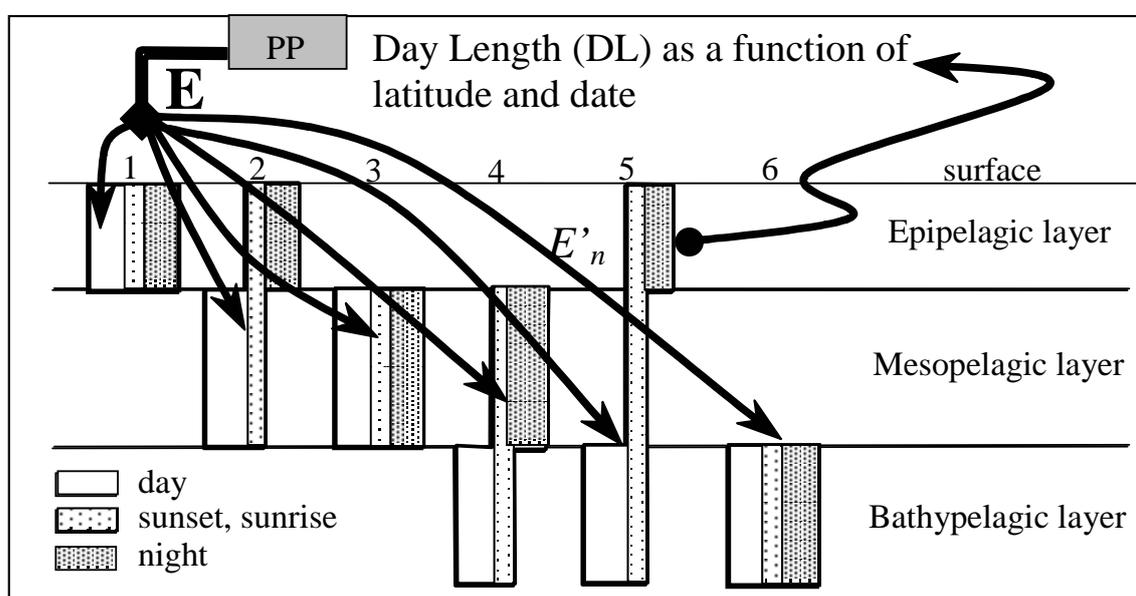
Figure 1.1: Daily vertical distribution patterns of the micronekton in the pelagic ecosystem: 1, epipelagic; 2, migrant mesopelagic; 3, non-migrant mesopelagic; 4, migrant bathy-pelagic; 5, highly-migrant bathypelagic; 6, non-migrant bathypelagic . The part of energy (E) transferred from primary production (PP) to intermediate trophic levels is redistributed (En') through the different components.
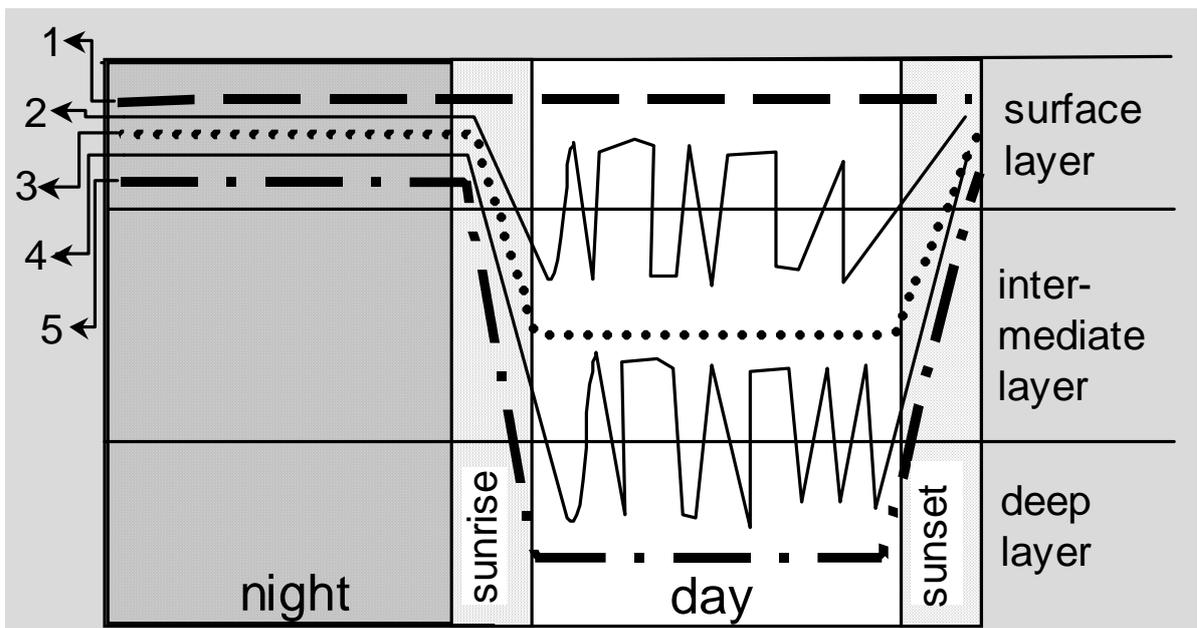
Figure 1.2: Five typical vertical movement behaviors simulated using a 3-layer and 2-type of prey pelagic system (adapted from Dagorn et al. 2000): 1- epipelagic predators (e.g., skipjack, marlins and sailfish); 2- predators moving between the surface and intermediate layers during the day (e.g., yellowfin tuna); 3- predators mainly in the intermediate layer during the day (e.g., albacore tuna); 4- predators moving between deep and intermediate layer during the day (e.g., blue shark); 5- predators mainly in the deep layer during the day (e.g., bigeye tuna and swordfish).

The simulations are driven by a bio-physical environment predicted from a coupled ocean physical-biogeochemical model. This new version of SEAPODYM includes expanded definitions of habitat indices, movements, and natural mortality based on empirical evidences. A thermal habitat of tuna species is derived from an individual heat budget model. The feeding habitat is computed according to the accessibility of tuna predator cohorts to the different vertically migrating and non-migrating micronekton (mid-trophic) functional groups. The spawning habitat is based on temperature and the coincidence of spawning fish with presence or absence of predators and food for larvae. The successful larval recruitment is linked to spawning stock biomass. Larvae drift with currents, while immature and adult tuna can move of their own volition, in addition to being advected by currents. A food requirement index is computed to adjust locally the natural mortality of cohorts based on food demand and accessibility to available forage components. Together these mechanisms induce bottom-up and top-down effects, and intra- (i.e. between cohorts) and inter-species interactions. The model is now operational for running multi-species, multi-fisheries simulations, and the structure of the model allows a validation from multiple data sources. However, running multi-species simulations in optimization mode requires very high computer power and has not been tested yet.

1.1.3   Model framework

The general scheme of the model and approach is presented on Fig. 3.1. The forcing for the model are predicted fields of physical environmental data (temperature and currents) and dissolved oxygen concentration, averaged by month over three depth layers: the surface epipelagic layer, the sub-surface mesopelagic layer and the deep bathypelagic layer. The definition of these layers is based on the euphotic depth (see Fig. 1.1). Primary production P is vertically integrated and is given in units of mmol C m^{-2}d^{-1}. These bio-physical environmental fields are used to describe the preferred habitat of the large pelagics for foraging and spawning and to predict their temporal and spatial dynamics. Anthropogenic forcing is represented by spatially-disaggregated fisheries effort data, grouped by gear type. Effort data are used to parameterize fishing mortality and therefore predicted catch.

1.2   Underlying equations

The population of tuna (or other pelagic top predator) in SEAPODYM is age- and stage-structured. Four life stages, larvae, juveniles and adults before and after maturity are modeled differently. Let k_s denotes the number of cohorts by stage and $\sum_4 k_s = K + 1$ - the total number of cohorts specified[1]. Let $N_a(t, x, y)$ denote the density of the cohort (age) $a = 0, .., K$ at time $t \in (t_0, t_n)$ and position $(x, y) \in \Omega \in \Re^2$.

---

[1]Note that index K depends on the maximal age of the population and the temporal discretization of the cohorts.

The micronekton organisms serve as a forage for modeled predator population. They are grouped into six classes $F_c(t, x, y), c = 1, .., 6$ (as shown on Fig. 1.1) by the common dial migration pattern. Since they are predicted by the separate unit of the model and used here a one of the forcing variables, we omit detailed description of the underlying model (see e.g. Lehodey, 2001).

Let us denote $\mathbf{v}_z = (u_z(t, x, y), v_z(t, x, y))$, the vector of oceanic horizontal currents, $T_z(t, x, y)$ is the temperature, and $O_z = O(t, x, y)$ is dissolved oxygen averaged through vertical layers $z = 0, 1, 2$. The definition of the layers descends from diurnal migrations exhibited by micronekton species through the water column (see Fig. 1.1) and their depths are linked to euphotic depth (see Section **??** for more details).

The transport of tuna population can be described by continuous advection-diffusion-reaction (ADR) equations describing. For brevity, we use gradient operator $\nabla = (\partial_x, \partial_y)^{\mathrm{T}}$, divergence operator of a vector field $\mathrm{div}(\mathbf{v}) = \partial_x u + \partial_y v$ and $\Delta = \mathrm{div}\ \mathrm{grad}$ for Laplacian of scalar field of population density. The state variables N_a as well as environmental forcing variables are determined at point (x, y) and time t (hereafter we will omit the notations of space and time).

The ADR system describing dynamics of age-structured tuna population is:

$$\partial_t N_a = \begin{cases} -\mathrm{div}(N_a \mathbf{v}_0) + \delta \Delta N_a - m_a N_a + S_a, \ a = 0, .., (k_L + k_J - 1); \\ -\mathrm{div}(N_a \tilde{\mathbf{v}} + N_a \mathbf{V}_a) + \mathrm{div}(D_a \nabla N_a) - (m_a + f_a)N_a + S_a, \ a = (k_L + k_J), ..., K. \end{cases}$$
$$(1.1)$$

where δ is constant diffusion coefficient of larvae and juveniles, $\tilde{\mathbf{v}}$ denotes weighted average (by the accessibility to depth layer, see section 1.3 for details) of oceanic currents through all layers, $\mathbf{V}_a$ is vector of directed velocity of adult tuna at age a, which is proportional to the gradient of the habitat index (see section 1.5), diffusion rates $D_a = D_a(x, y) = D(a, T, O, F)$ are functions of age and environmental factors, $m_a = m_a(x, y)$ are natural mortality and f_a are fishing mortality by adult cohorts (mortality functions are detailed in 1.6). Terms S_a represent sources of the new population density to corresponding variable and include both survivals from younger age classes as well as the effects of spawning and recruitment.

The system (1.1) is completed by Neumann boundary conditions describing imperme-ability of the domain bounds $\partial\Omega$:

$$\mathbf{n} \cdot \mathbf{v}\Big|_{\mathbf{x} \in \partial\Omega} = \mathbf{n} \cdot \nabla N_a \Big|_{\mathbf{x} \in \partial\Omega} = 0 \qquad (1.2)$$

According to these conditions no additional source of biomass and no loss are possible within the domain when recruitment and mortality are absent. It implies the requirement on the domain definition to be naturally closed for considered tuna species.

Since age discretization in time for most of the cohorts is different from the time step chosen to numerically solve the system (1.1-1.2), ageing equations are added in order to imply transition from one age group to another. Hence, the system of continuous ADR equations

is supplemented by the discrete equations describing the following. At each time step tuna cohort comprises a number of individuals remaining in the age class at a current time step plus recruits from younger age, minus the number of individuals which pass to the older age group. Thus, we have simple relationships:

$$N_a^{t+1} = p_{n,a-1}N_{a-1}^t + (1 - p_{n,a})N_a^t, \ a = 1, ..., K \tag{1.3}$$

The survival coefficients $p_{n,a}$ determine the rates of decay of the density due to natural, predation and fishing mortality depending on the time spent in corresponding age. They are relative values between 0 and 1, such that

$$p_{n,a} = \frac{e^{-n(m_a+f_a)}}{\sum\limits_{i=1}^{n} e^{-i(m_a+f_a)}} \tag{1.4}$$

where n is the ratio between time step in population age structure and the time step of discretization in numerical approximation of (1.1-1.3), i.e., $\varDelta a = n \varDelta t$(month). Note that if $n = 1$, i.e., age discretization coincides with time discretization, then $p_{1,a} = 1$ and the corresponding equation, i.e., Eq. 1.4 simplifies to $N_1^{t+1} = N_0^t$. Note also that the fishing mortality $f_a = 0$ for unexploited part of population.

The new spawning density of the tuna population is the product of two functions, the Beverton-Holt relationship giving the dependence on the density of mature adult tuna and H_S, the spawning habitat index (see Eq. 2.10):

$$S_0 = \frac{RN}{1 + bN} \cdot H_S, \tag{1.5}$$

Note that setting parameter $b = 0$ (in parameter file, see chapter 2) gives Malthusian growth of population density although still restricted by the habitat conditions.

1.3 Habitats and indices

Physical and biogeochemical conditions influence fish population dynamics through changes in spawning conditions, habitat suitability, and distributions of food resources, thus inducing changes in fish movement behavior, reproduction and mortality. Environmental data are used in SEAPODYM to functionally characterize habitat of the population depending on its thermal biochemical and forage preferences. Generally there are three types of habitat in the model: 1) thermal, 2) spawning and 3) feeding. The thermal habitat is deduced using a size/weight-based relationship by age. The spawning habitat depends on thermal habitat and includes also the food and predation factor. Feeding habitat is proportional to the prey

abundance with the proportionality coefficients being accessibility to the layer where the prey components live. Using these definitions we construct the habitat suitability indices for each cohort as the habitat preferences vary with age. These indices are then used to control population dynamical processes (both spatial and temporal) such as movement to the feeding or spawning grounds, natural mortality and predation.

1.3.1   Fish thermal habitat

At the scale of a population, we consider that the thermal habitat can be represented by a Gaussian distribution with an average temperature linked to the size-dependent body temperature at steady state, and by a standard error of the distribution linked to the thermal inertia of the fish. Here, we assume that for a given species, there is an optimal intrinsic temperature (T_i^*) that remains constant, whatever the age/size, and that this temperature is a target temperature for any individual of the species (e.g., due to genetic and physiological adaptation during species evolution). It follows from a tuna heat budget model (Holland et al., 1992) that when becoming larger, the fish will have to search for colder habitat temperature to compensate their increasing body temperature at steady state. But they will have also larger temperature range due to their increasing with size l thermal inertia. Therefore we define the population size-dependent thermal habitat by a Gaussian distribution $N(T^*(l), \sigma(l))$:

$$H_T(l) = e^{-\frac{(T - T^*(l))^2}{2\sigma^2(l)}} \tag{1.6}$$

with linearly decreasing with size function $T^*(l)$ and linearly increasing with weight function $\sigma(l)$ (Fig. 1.3).

Let each cohort a be defined by the mean fork-length at age l_a and mean weight at age w_a. Considering the age discretization of the population our definition of the species thermal habitat implies that the maximal average temperature occurs at age 0, i.e., spawning. We assume that the species intrinsic temperature T_i^* is equivalent to the average temperature of the first cohort, i.e., that it is defined as an optimal spawning temperature $(T_i^* = T_0^*)$. Hence, for a given cohort $a \in (0, K)$, the optimal temperature is:

$$T_a^* = T^*(l_a) = T_0^* + (T_K^* - T_0^*)\frac{l_a}{l_K} \tag{1.7}$$

and the associated std. dev.:

$$\sigma_a = \sigma_0 + (\sigma_K - \sigma_0)\frac{w_a}{w_K} \tag{1.8}$$
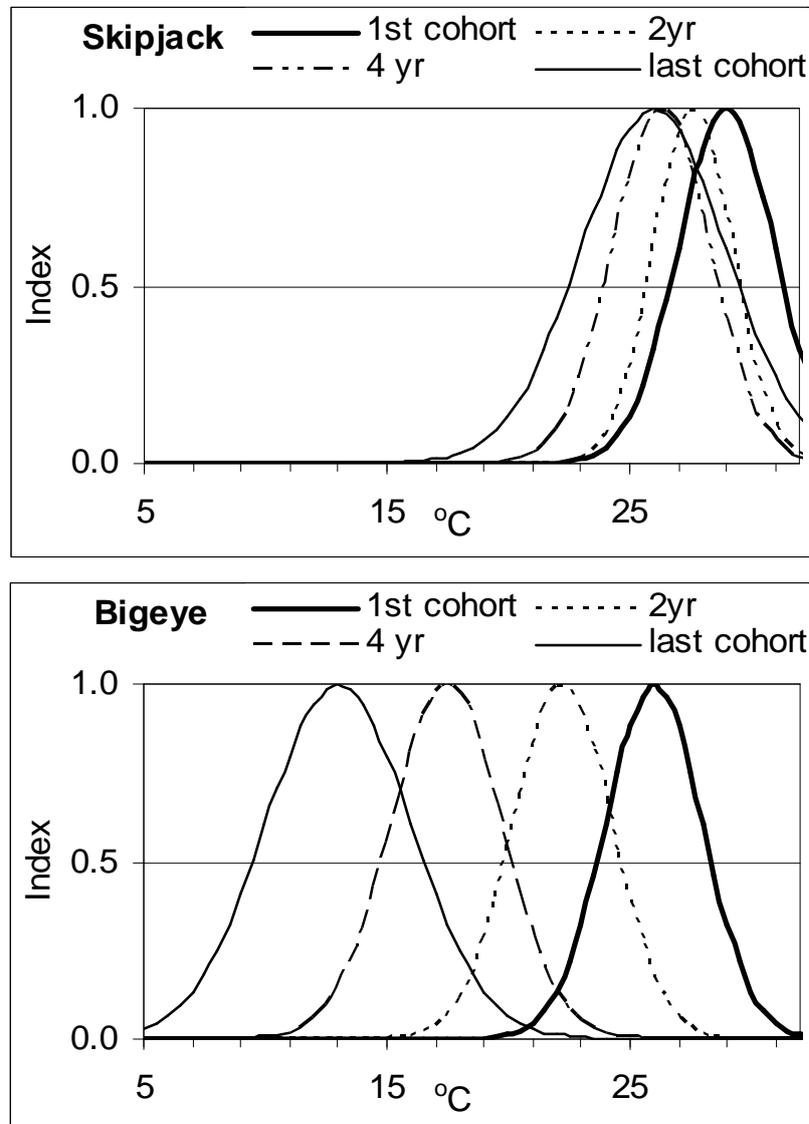
Figure 1.3: Change in temperature function used to define the habitat temperature of the population based on a Gaussian distribution with linearly decreasing with size average temperature and increasing with weight standard deviation. Example of parameterization used for skipjack and bigeye. The average size of the latest cohort for these two species is 80.15 cm and 175.08 cm respectively, with corresponding weights of 10.61 kg and 113.93 kg. Redrawn from Lehodey et al. (2008).

However, for practical reason to facilitate parameter estimation and also due to usually low sensitivity of the model to σ_0 parameter (in the absence of data on the larvae distribution), the last equation has been modified to:

$$\sigma_a = \sigma_T + \delta \frac{w_a}{w_K} \tag{1.9}$$

where σ_T is allowed to be estimated during function minimization procedure while δ can be fixed at different values (can be added later to the list of control parameters as well).

1.3.2   Spawning habitat

The spawning habitat H_S is used to constrain larval production and mortality of larvae cohort(s). This index and larvae dynamics are defined to represent four mechanisms that are assumed to control the survival of larvae and the subsequent recruitment:

- changes in the spatial extent of the spawning habitat with temperature;

- coincidence of spawning with presence or absence of food for larvae (micro-zooplankton, that is approximated by primary production), i.e., the match/mismatch mechanism proposed by Cushing (1975);

- coincidence of spawning with presence or absence of predators of larvae (that are the micronektonic organisms, i.e., the prey of adults);

- redistribution of larvae by the oceanic circulation that can create retention of larvae in favorable areas with lower natural mortality or conversely move the larvae in unfavorable zones where the natural mortality will be higher.

Therefore, spawning habitat index is a function of surface layer temperature T_0, micronekton biomass in the surface layer F_0 and primary production converted to the wet weight of zooplankton P_{ww}. So, if ocean primary production P is given in mmolC/m^2, after conversion to the wet weight of zooplankton species the units of P_{ww} become g/m^2.

Since larvae predation is likely maximal during daytime and twilight periods, the biomass of predator of larvae F_0 is computed as the sum of mid-trophic biomass in the epipelagic layer during daytime (the daylength) and epi- and migrant-pelagic mid-trophic biomass during the sunrise and sunset periods. Practically, a period of 2 hours is used, i.e., 2/24th of the night-time epipelagic biomass.

Let $\Lambda = EP_{ww}/F_0$ denote the ratio between food for larvae and the tuna forage that is considered as the potential predator for larvae. The constant E is the energy transfer coefficient.

The spawning habitat is described by the following function:

$$H_S = \phi(\Lambda) \cdot H_{T_0},$$

where $\phi(\Lambda)$ is the nonlinear saturation function determined in [0,1] interval:

$$\phi(\Lambda) = \frac{\Lambda}{\alpha + \Lambda} \tag{1.10}$$

The curvature parameter α is unknown and included in the list of parameters to be estimated from the data. The amplitude of the trade-off effect between food and predator of larvae increases with α. If $\alpha = 0$, only temperature has an effect. The dependence on temperature is described by a Gaussian function as detailed in section **2.3.1** above:

$$H_{T_0} = e^{-\frac{(T_0 - T_0^*)^2}{2\sigma_0^2}} \tag{1.11}$$

where the parameters T_0^* and σ_0 are the optimal temperature and width of tolerance interval (standard deviation) in the Gaussian for the larvae stage.

1.3.3 Feeding habitat

The feeding habitat index expresses how a particular place and time is favorable for feeding on prey components, based on the accessibility to each of them considering both the physical ability of the predator (different by species and by age) and the physical conditions in the water layer inhabited by the forage component at different time of the day. Thus, although SEAPODYM model is two-dimensional in space, the third dimension is included in feeding habitat through vertical distribution of the forage components, i.e at each spatial position (x, y) the habitat H_F is the sum of abundances of preys living in the water column weighted by accessibility coefficients to each z depth layer:

$$H_F = \sum_z \Theta_z \left(F_{zz} + \tau \sum_{k \neq z} F_{zk} + (1 - \tau) \sum_{k \neq z} F_{kz} \right) \tag{1.12}$$

where τ denotes the fraction (time) of the daylight in a day, and F_{zz} are the elements of the forage matrix with diagonal elements being z-resident (non-migrant) forage class, and off-diagonal elements are the migrants (as defined in **2.1.2**, see also Fig. **1.1**). The first index shows the layer the forage stays during the day, the second denotes for the night layer correspondingly:

$$\mathbf{F} = \begin{pmatrix} F_{11} & 0 & 0 \\ F_{21} & F_{22} & 0 \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

The accessibilities Θ_z include certain environmental conditions as limiting factors to access the layer. Two factors, temperature and oxygen, are considered to be essential for defining the extents of tuna's vertical habitat and hence the quality of the habitat in terms

of amount of food available. It is known that low levels of oxygen make the tuna habitat unfavorable (see, e.g. Brill, 1994), so the influence of the oxygen on the habitat quality can be described by a sigmoid function:

$$\Psi(O_z) = \frac{1}{1 + e^{\gamma(O_z - \hat{O})}}$$

with the critical value $\hat{O}$ being variable for different species (see Fig. 1.4).

The temperature factor is included in the form of thermal habitat (see section **2.3.1**), hence making the accessibility coefficient dependent on age:

$$\Theta_{z,a} = \Psi(O_z) \cdot H_{T,a} \tag{1.13}$$

Thus, the feeding habitat index $H_{F,a}$ is computed for all adult cohorts, at both immature (young, $a = k_L + k_J + 1, .., k_Y$) and mature ($a = k_Y + 1, .., K$) stages. This habitat plays main role in the movements of the cohorts, representing the displacement toward favorable feeding grounds and is used to define movement index (see **Movement dynamics** section).
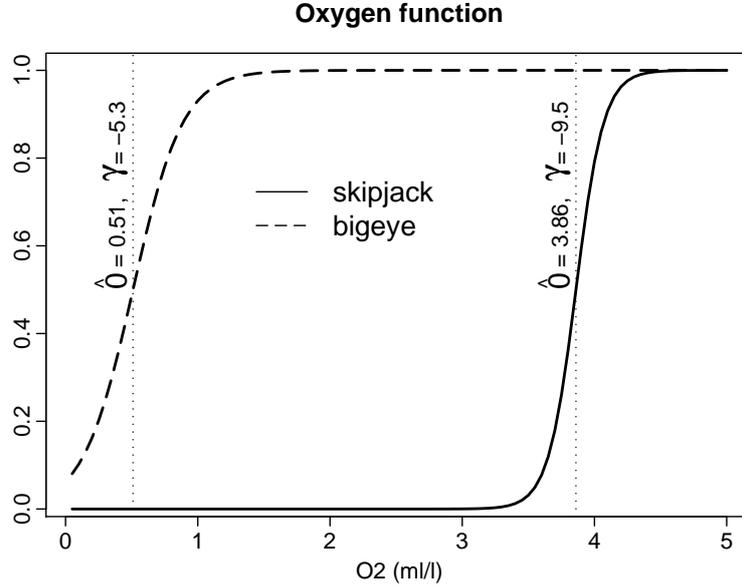

Figure 1.4: Dependence of feeding habitat on oxygen level. Parameters for skipjack and bigeye tunas were obtained with help of optimization procedure using catch and catch at size data.

Predator's diet. The definition of the habitat index allows us to compute the diet of the predators (by age, time and space). Note that Eq. 1.12 can be rewritten as $H_{F,a} = \sum_{dn} F_{dn} (\tau \Theta_{d,a} + (1 - \tau)\Theta_{n,a})$, i.e. as the sum forage components weighted by their average (over day and night) accessibilities to tunas. This enables computing the relative accessibilities. Let's denote the accessibility to forage component (the expression in parenthesis) as $\Theta_{c,a}$ i.e.,

$$\vartheta_{c,a} = \frac{\Theta_{c,a}}{\sum_n \Theta_{c,a}} \tag{1.14}$$

can be described as the time the predator spends searching for the n-th forage component. Hence, the product $\vartheta_{c,a}F_c$ represents the portion of the c-th forage in the predator's diet. This approach is illustrated on Fig 1.5 for two tuna species - Pacific skipjack *Katsuwonus pelamis* and bigeye *Thunnus obesus*. Given the level of potential details predicted by the model, data were aggregated through a few regions and averaged through all young ($a = k_L + k_J + 1, .., k_Y$) and adult ($a = k_Y + 1, .., K$) cohorts. Overall, there are limited changes for skipjack considering regions or life stages. This species is mainly limited to the epipelagic layer by its thermal habitat and consequently its diet is dominated by epipelagic group ($\approx 2/3$). For bigeye, the diet is more variable between young and adult fish and regions. The diet of young fish is dominated by non-migrant mesopelagic species in the west, by migrant-mesopelagic in the north-east, and by equal proportions of both migrant and non-migrant mesopelagic species in the east. When they are adult, bigeye have a larger thermal range and can access to all forage components except in the coldest northeast region where they cannot access the meso- and bathy-pelagic layer.

1.4   Habitat indices.

To avoid unexpected dynamical behavior of the model, for example obtaining negative diffusion parameters, we need to constrain the habitat indices, in other words, scale them in the bounded space. For simplicity we scale them within $[0, 1]$ segment. As it is seen from Eqs. 1.10 and 1.11 the spawning index is scaled by default. To constrain the feeding and consequently the movement index we use trigonometric scaling:

1.5   Movement dynamics

The model simulates passive transport of tuna larvae and juveniles by oceanic currents in the surface (epipelagic) layer, while young and adult tuna can direct their own movements and swim at a speed proportional to their size and the gradient of adult habitat index, in addition to being advected by currents.
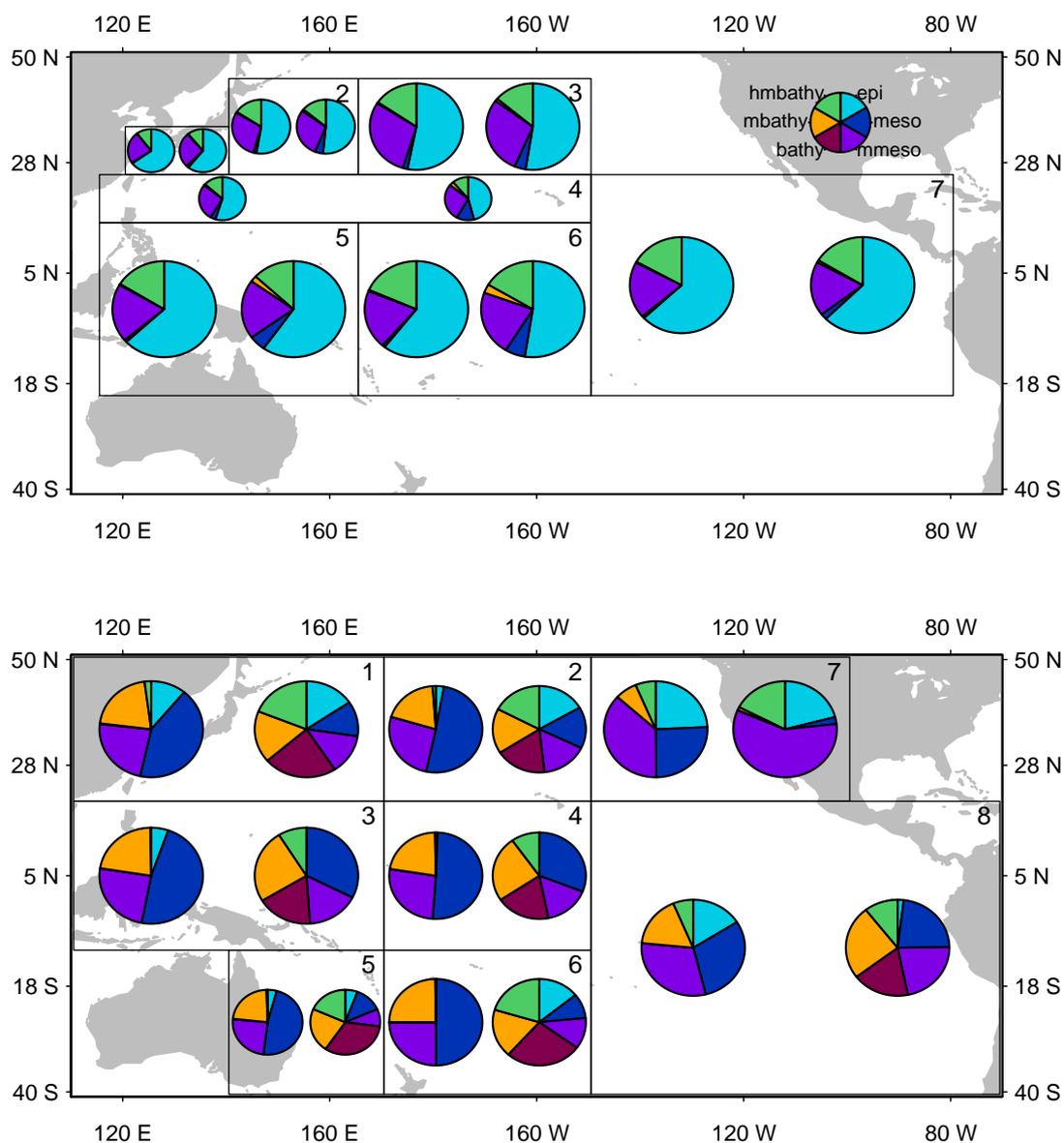
Figure 1.5: Diet compositions of skipjack (top) and bigeye (bottom). Predictions are aggregated over 10 years (1995-2004) in sub-regions following a stratification based on fishing data used for stock assessment analyses (Hampton et al 2006). There are two pie-plots by region for young immature (left) and adult mature (right) fish.

These movements are modelled with two components, namely random dispersal and directed displacements, described by diffusion and advective term in the equations (2) respectively. For both types of movements, the displacement per time unit is obviously directly dependent on the size of the individuals. Therefore, the movement is linked to a maximum sustainable speed $V_{\max,a}$ expressed in Body Lengths per second.

As pointed out earlier, in the case of larvae and juveniles, the advection is due to oceanic currents only. An 'age of autonomous movement' is defined from which fish (cohorts) can swim by their own volition (if set to 0, the first cohort with autonomous movements is the first one following the juvenile phase). In this latter case, as in conventional chemotaxis models (see Keller and Segel, 1971, Czaran, 1998, Turchin, 1998) velocity field ($\mathbf{V}_a$) is proportional to the gradient of the movement stimuli, represented here by the quality of adult habitat.

$$\mathbf{V}_a = \chi_a \left(\frac{\partial I_{2,a}}{\partial x}, \frac{\partial I_{2,a}}{\partial y} \right)^{\mathrm{T}} \tag{1.15}$$

where the taxis activity constant χ_a is proportional to $V_{\max,a}$ which is, in turn, inversely related to the average size at age (Malte *et al.*, 2004) following $V_{\max,a} = V_{\mathrm{m}} \left(1 - \eta \frac{l}{l_K} \right)$, where the parameter $\eta = 0.1$ implies small negative slope.

In addition, we can expect that individuals will tend to stay longer in the presence of favorable conditions (low diffusion) but will want to escape quickly from unfavorable habitats (high diffusion), so that diffusion values should be linked to habitat condition. We define maximal diffusion coefficient in the null (extremely unfavorable) habitat according to the formula of two-dimensional mean square dispersal (see, e.g. Turchin, 1998), namely $D_{\max} = R^2/4t$, or if we assume that during time t individual will cover the maximal distance moving with its maximal sustainable speed $V_{\max}$, we have as upper estimate of diffusion coefficient $D_{\max} = V_{\max}^2 t/4$. Thus, in each habitat, a given upper value is reduced according to non-linear relationship with the habitat index $I_{2,a}$ and linear relationship with its gradient, $\nabla I_{2,a}$:

$$D_a = D_{\max} \left(1 - \frac{I_{2,a}}{c + I_{2,a}} \right) (1 - \rho \left| \nabla I_{2,a} \right|) \tag{1.16}$$

where c is the coefficient of variability of fish diffusion rate with habitat index. The expression $(1 - \rho \left| \nabla I_{2,a} \right|)$ with $\rho < 1$ balances diffusive and advective movements to ensure that maximal displacement due to both diffusion and taxis does not exceed the distance which fish can cover with its maximal sustainable speed.

Movement index The movement index, denoted I_a (previously called adult habitat index) must navigate the cohorts throughout the year, i.e. has to include seasonality of the migrations for those cohorts which exhibit such behavior due to environmental variability.

Thus, while for young immature cohorts the habitat can be described by the feeding index, it must enable the movement of mature fish toward favorable spawning grounds due to the changes of daylight length. The gradient of the day length $\partial_t d$ can serve as a trigger switching tuna behavior from foraging to searching for spawning grounds. One of the hypothesis of how this search occurs assumes that adult tuna tend to direct their movements to find a place with environmental conditions as those occurring during their birth (see e.g., Cury, 1994), i.e., in the model the conditions defined by the spawning habitat index. Based on such assumption, we reproduce the seasonal change between feeding and spawning migrations by switching the movement index that drives the movement of fish from the feeding habitat to the spawning habitat using a threshold in the rate of increase of the day length as a triggering effect. The equation used for $a = k_Y + 1, .., K$, is:

$$I_a = \frac{H_{F,a}}{1 + e^{\kappa(\partial_t d - \hat{G})}} + \frac{H_S}{1 + e^{\kappa(\hat{G} - \partial_t d)}}, \tag{1.17}$$

where $\hat{G}$ is a triggering value of the daylight gradient and κ is large constant $(\approx 10^3)$ producing abrupt but continuous shift between feeding and spawning indexes.

This switch occurs earlier for fish that are in higher latitudes, and therefore far from their spawning grounds in warm waters (Fig. 1.6). Note that in low latitudes, the gradient of the day-length is too low to have any effect and adult habitat is always driven by the feeding habitat, leading in that case to opportunistic reproduction.

used to constrain both diffusion and advection terms in the modeling of fish movements and

1.6 Mortality

$m_0 = m_0(x, y) = f(T_0, P, F)$ is larval (age zero) natural mortality rate, a function of water temperature at surface layer T_0, primary production P and forage density F; $m_{1,2} = m_{1,2}(x, y) = f(T_0, N)$ are juvenile (ages 1, 2) natural mortality rates dependent on surface layer temperature and total adult tuna density.

Average natural mortality by age (in months, τ) in the population is described as the sum of two functions. An exponentially decreasing function with age represents the mortality during the early life history (mainly due to starvation and predation). An increasing function with age characterizes natural mortality in the adult phase, i.e., mainly senescence and diseases:

$$m_S(\tau) = \bar{m}_S \left(1 + e^{\beta_S(\tau - A)}\right)^{-1} \tag{1.18}$$

$$m_P(\tau) = \bar{m}_P e^{-\tau \beta_P} \tag{1.19}$$
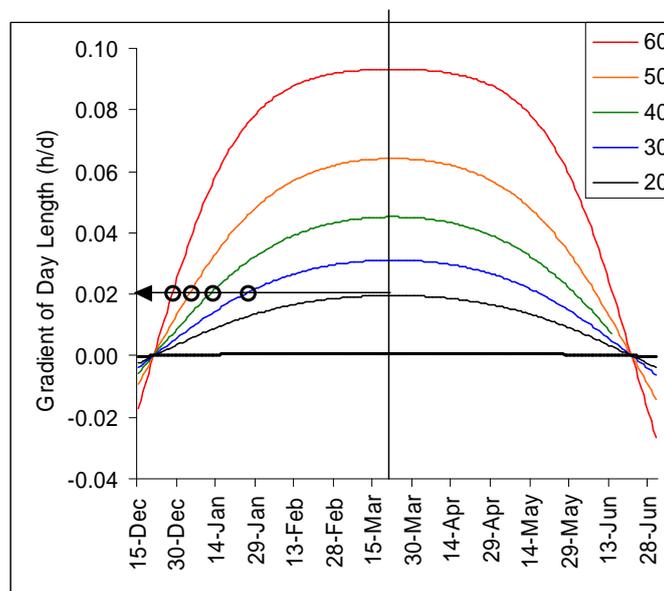
Figure 1.6: Example of seasonal cycles in gradient of day length at different latitudes in the North hemisphere. With a threshold set above 0.02 hours per day, the adult habitat that constrains the movement of fish will switch from feeding habitat to spawning habitat at the end of December for mature fish at latitude 60, and one month later at latitude 30. Movement of fish at latitude below 20 remains always driven by the feeding habitat.

where $\bar{m}_S$ and $\bar{m}_P$ are maximal senescence and predation mortalities, β_S and β_P are slope coefficients, and A is the age at which $m_S(A) = \bar{m}_S/2$. The sum of the two components gives the total natural mortality-at-age coefficient:

$$m(\tau) = m_S(\tau) + m_P(\tau) \tag{1.20}$$

To this average coefficient-at-age, it is possible to add local variability linked to environmental conditions, expressed by any index I standardized between 0 and 1, e.g.:

$$M = m(\tau)(1 - I + \varepsilon) \tag{1.21}$$

For larvae, the spawning habitat index is used, while two other indices are defined for juvenile and adult stages.

Juvenile index. The juvenile index I_J combines the thermal habitat function and a function which accounts for predation by adults, either from the same species (cannibalism) or other tuna species (in the case of multi species simulations):

$$I_{JP} = \left(1 - \frac{\langle N \rangle}{h + \langle N \rangle}\right) \cdot \Phi(T_0) \tag{1.22}$$

where $\langle N \rangle = \sum_{a=1}^{K} N_a$, i.e., is the total size of adult portion of the population, and the unknown parameter h determines the predation intensity in the habitat depending on the total number of adults tuna being present locally.

Food Requirement index. Though the feeding habitat is used to constrain the movement of fish there is not a direct link between the available biomass of prey (mid-trophic groups) and the food requirement of the predators (tuna). In this version of the model, a food requirement index $I_{FR,a}$ is used to adjust locally the natural mortality of the species, based on intra- and inter-species competition for food. We define the adult food requirement index as the ratio between available forage in the habitat and food required by adult tuna at age a:

$$I_{FR,a} = \frac{\sum_n F_n}{\psi \sum_n r w_a N_a \vartheta_{a,n}}$$

where r is the food ration of an individual, i.e., proportion of tuna weight w_a, ψ is a parameter responsible for consumption of forage by other predators and ϑ is the relative accessibility coefficient (see Eq. 1.14), i.e.,
Finally, in order to scale this index between 0 and 1 we use the transformation:

$$I_{FR,a} = \frac{1}{1 + I_{FR,a}^{\nu}} \tag{1.23}$$

Thus, this index will vary from 0 when the requirement by the cohort is a high contribution to a total forage requirement that is much higher than available forage, to 1 when this requirement is a low contribution to a total forage requirement that is much lower than available forage.

An application with two tuna species (skipjack and bigeye) is used to illustrate how this index provides a relative measure of food competition between cohorts of all predator species explicitly represented in the model (1.7). The parameter (epsilon) was set to 0.5, which means that mortality defined by equation above is lower when food requirement index I_{FR} is above 0.5 and higher when I_{FR} is below 0.5.

For skipjack, there is a slightly lower biomass in the case of the two-species simulation in both WCPO and EPO. For bigeye however, the decrease is much higher with a 57 % biomass reduction. The very large biomass of skipjack has logically a strong impact on bigeye cohorts through food competition, especially in the WCPO which is the core habitat of skipjack. When computing the mortality coefficients over the two east and west regions by age resulting from mortality equation, we obtain effectively higher average mortality rates in the case of the multi-species simulation. But these curves also indicate that this interaction occurs more strongly between the oldest cohorts of skipjack and the youngest cohorts of bigeye. In growing older, bigeye access deep forage components that are not available for skipjack (Fig. 1.5) and thus the mortality effect due to food competition gradually decreases.

Fishing mortality. The model includes a description of multiple fisheries. For each fishery, the actual fishing effort is used to predict catch of tuna species by cell, time step and by age class of species. The fishing mortality, for a given cell and at a given time step, is proportional to the sum of fishing efforts in the cell, the catchability coefficients of the corresponding fisheries and the selectivity coefficients for the gear and age (size) considered. The fishing mortality (f), added to the natural mortality (M), gives the total mortality rate (Z).

Each fishery has a single catchability coefficient by species. Catchability coefficients are corrected to the surface of the cell. Since this coefficient is constant it is recommended to define homogeneous fisheries. A first obvious criterion for example is the fishing gear, but it may be also necessary to consider some large spatial-scale stratification, different fishing strategies, or different periods of time associated to large changes in the fishing strategy or gear technology.

$$f_a = sum_f s_{f,a} q_f E_f, \ \ a = 1, ..., K. \tag{1.24}$$
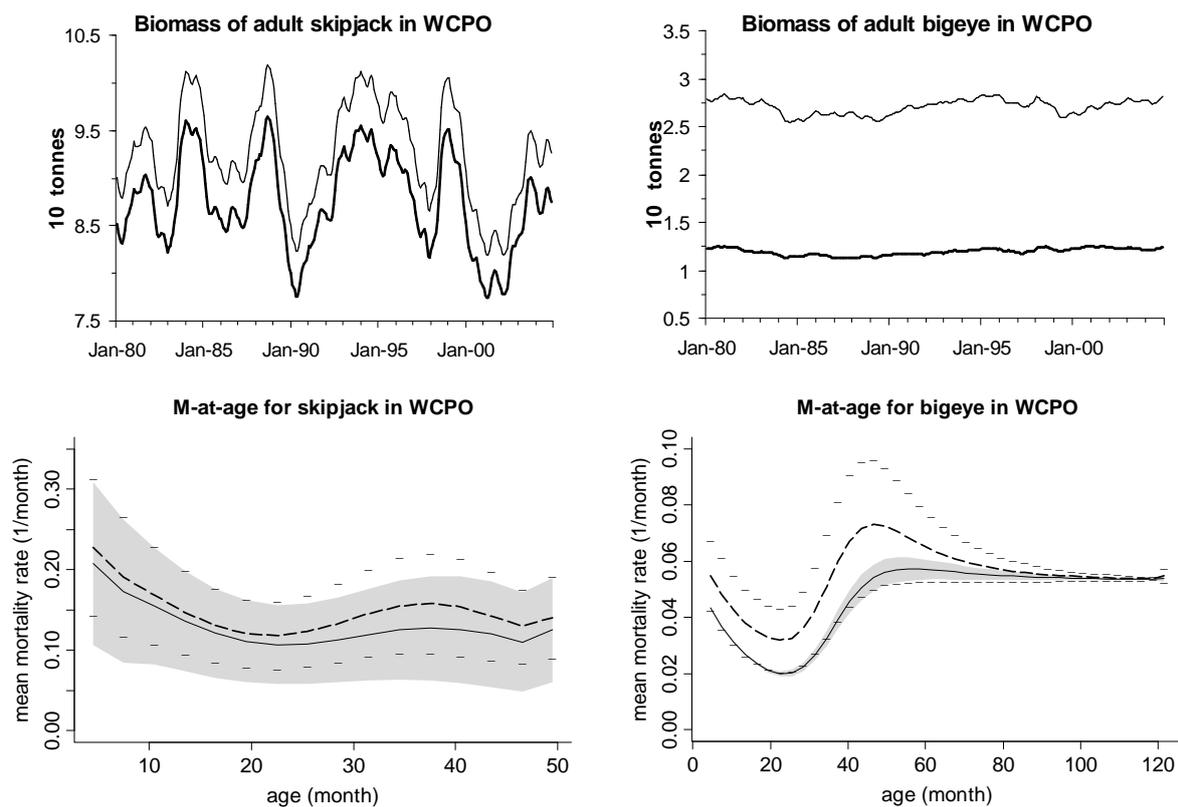
Figure 1.7: Time series of biomass of adult skipjack and bigeye comparing single (thin curves) vs. multiple (thick curves) species simulations in the western-central (WCPO) Pacific Ocean (i.e., west of 150W), and corresponding average and standard errors of mortality by age modified by the food requirement index I_{FR} for skipjack and bigeye. Thin curves and grey shaded area are for single species simulation and dotted curves are for multi-species simulation.

where q_f is the catchability coefficient of fishery f, E_f is observed fishing effort and $s_{f,a}$ is fishery-specific selectivity, which is specified as either sigmoid function (type I selectivity function) of age or asymmetric Gaussian (type II):

$$s_{f,a} = \begin{cases} \left(1 + e^{-\varsigma_f(l_a - \hat{l}_f)}\right)^{-1}, & \text{type I} \\ e^{-\frac{(l_a - \hat{l}_f)^2}{\sigma_{s_f}}}, & \text{if } l_a \leq \hat{l}, \text{ type II}, \\ \mu_f + (1 - \mu_f)e^{-\frac{(l_a - \hat{l}_f)^2}{\sigma_{s_f}}}, & \text{if } l_a > \hat{l}, \text{ type II}, \end{cases} \tag{1.25}$$

Total mortality by life stage. Mortalities of larvae and juveniles can vary locally around the average coefficient-at-age value, i.e., if $I < 0.5$ mortality rate increases and the opposite is true for $I > 0.5$. Adult mortality, in contrast, can only increase depending on the food requirement index $I_{FR,a}$ that determines the level of food deficit for each age group. Such penalty leads to highest local mortality rates for young tunas. The coefficient q_f is catchability of fishery f, E_f is observed fishing effort and $s_{f,a}$ is fleet-specific selectivity, which is specified as either sigmoid function (type I selectivity function) of age or asymmetric Gaussian (type II):

With added effects of fishing and environmental variability expressed through habitat index functions, the local mortality rates of each cohort are:

$$m_0 = M(\tau_0)(1 - I_0 + \varepsilon) \tag{1.26}$$

$$m_k = M(\tau_k)(1 - I_1 + \varepsilon), \ k = 1, 2 \tag{1.27}$$

$$M_a = M(\tau_a)\left(1 + e^{2I_{FR,a}-1}\right) + \sum_f s_{f,a}q_f E_f, \ a = 1, ..., K. \tag{1.28}$$

Mortalities of larvae and juveniles can vary in both directions, i.e., if $I < 0.5$ mortality rate increases and the opposite is true for $I > 0.5$. Adult mortality, in contrast, can only increase depending on the food requirement index $I_{FR,a}$ that determines the level of food deficit for each age group. Such penalty leads to highest local mortality rates for young tunas.

Chapter 2

Running the model

2.1 Getting started

System requirements SEAPODYM is a Linux application however it is possible also to compile and run it under Windows with Cygwin installed. When application is used for simulation runs only, namely, for predicting population dynamics with fixed (presumably best known) parameters, there are no special requirements for the computer configuration. However, if parameter estimation is intended, to run the application efficiently, one needs to have an access to a machine with 64-bit CPU 2GHz and higher, and at least 8Gb of RAM on board. The latter is the key requirement because the backward (adjoint) differentiation method stores all intermediate variables needed for exact evaluation of a cost function gradient in the operational memory. In case if there is not enough RAM available, the program will dump all temporary data on a hard disk (in files cmpdiff.tmp and gradfil.tmp) that can significantly slow down the overall parameter estimation process.

2.1.1 Installations

This manual is for users who want to run application with the code and executable of the program that are provided. If modification of some part of the model code is desirable it is advised to contact the developer's team since changes in the code can interfere with computation of the gradient. Compilation of the source code requires to install two additional libraries listed below.

Autodif The version of ADModel Builder[1] pre-compiled with gcc3.4.6 for Linux 64-bit architecture machine. It can be downloaded from the following address:
http://code.google.com/p/admb-project/downloads/list

Simply uncompress downloaded archive to your local disk and specify the proper path in Makefile.

---

[1]Note, the program will not use ADModel Builder itself, but only Autodif libraries libado and libadt.

libxml2 To edit and store all model parameters as well as program configuration we use text-based **xml** parameter file. Reading this files requires libxml2 library. If not installed you can download and build library from address
ftp://xmlsoft.org/libxml2

After libraries are installed, create the source directory, place the code there, edit provided Makefile.i64 (see Appendix B) to be sure that you have correct paths, then compile SEAPODYM application typing the command:

```
make -f Makefile.i64
```

If there are no conflicts in libraries and compiler versions, all paths are specified correctly, the executable **seapodym_coupled64** will be built in the working directory. Now you are ready to setup the **running** directory, prepare the input forcing files, fishing data and configure the model and application parameters.

2.1.2   Running directory

We recommend the following structure and naming for your running directory:

```
/run-OGCMname-domain-resolution/
./fisheries/species-name/
./forage/
./init/
./mask.txt
./topo.txt
./OGCM-data DYM files
```

i.e. all fishing data files (see 2.1.4) will stored in the **fisheries** directory, **forage** directory will contain predicted micronekton distributions, in the folder **init** you will store the files with initial distributions (2.2), mask.txt and topo.txt are the files with land mask and topographic indices (2.1.3), the rest are DYM files with physical and biogeochemical forcing (2.1.4).

2.1.3   Defining the domain of simulation

The first step in preparing the model simulation is to define the model domain, which is the rectangular geographic area assumed to cover the habitat of the modeled species. The domain is discretized with help of a grid, with the size of cells reasonable to resolve studied dynamical features. To build the grid for the chosen geographic area we suggest to use the software called GMB. Using ETOPO2 (2-minute gridded global relief data, see information on http:

//www.ngdc.noaa.gov/mgg/global) topography maps GMB allows building regular and stretched grids and creating land masks and masks for arbitrary chosen depth level. For more details on how to use GMB see Readme.txt file that is provided with .jar executable. Notice that for the moment we consider regular grids only, i.e., orthogonal grids with constant resolution in latitude and longitude, the mixed-resolution grids implementation is envisaged for the next release of the SEAPODYM application with parameter estimation.

Below is an example of domain, grid and mask configuration, which was used in the analysis of Pacific skipjack population dynamics. This species is highly migratory and inhabits Pacific oceanic waters from West to East and appears in longline catches in higher latitudes up to 45N and 40S. The model of micronekton populations was run on a more extended toward higher latitudes domain, that is why the chosen coordinates (see parfile in Appendix C) are:

```
latitudeMin value=-54.0     most southern latitude (from +90N to -90S)
latitudeMax value=64.0      most northern latitude (from +90N to -90S)
longitudeMin value=90.5     most western longitude (from 0 to 360)
longitudeMax value=291.5    most eastern longitude (from 0 to 360)
nb_layer= 3                 number of vertical layers
deltaX value=120            resolution in nautical miles on the x (latitude) axis
deltaY value=120            resolution in nautical miles on the y (longitude) axis
```

Now, using GMB extract this area with ETOPO topographic data (see Fig **??**), build the two-degree grid and a create a land mask (GMB user's menu activates successively: 1) load map; 2) build grid; 3) create mask, etc.). Note that land mask created from topography data can be different from the one being used in OGCM or NPZD models, that is why it is recommended to use the data mask (mask, extracted from oceanographic and biogeochemical data) as initial one (use menu File/Load mask..). Addition of two more layer masks, i.e. with the depth defining epipelagic and mesopelagic layers, can be done in GMB as well (use menu Mask/Mask properties.. to set the depth and mask flag, then clicking "'create mask"' will add another layer with specified properties) in case if depth of the layer is considered to be constant, however, for the definition of variable layer depths a bit more complex procedure must be taken (see below). Important step to be taken before saving the mask file is to mask existing near-border cells which will cause the biomass leak in the finite-difference approximation scheme of ADR equation. Leak occurs when one computational cell is surrounded by three land cells (or if the ocean cells create one-cell channel), so it makes it problematic to create reflecting boundary condition. The GMB software filters such boundary cells in one click. Finally, the mask is saved by GMB into simple text file with the table of numbers, masking land by 0 and ocean (computational) cells by 1, 2 and 3 etc. according to the number of vertical layers specified for the simulation (see Fig. reffig:domain).

As described in Chapter 1 (see 1.1.1), in the model of micronekton we define six mid-

trophic classes distinguished by their diel migration pattern, which provides vertical strati-fication of these species habitat. It is assumed that euphotic depth z_{eu} (the depth at which light intensity falls to 1% of the value at the surface, for the method of how to calculate it see, for example, Morel and Berthon, 1989 is crucial in such stratification. Hence, the integration of physical and biogeochemical data must be done using euphotic depth (see next section), but also, to avoid changes in the borders of computational area for deeper layers (due to dynamic nature of euphotic depth) a 3-layer mask with fixed depths is computed as follows:

- The distribution of average and standard deviation of Z_{eu} is computed for all the time series

- A first test is used to set Z_{eu} to the mean + 2 std. dev. if Z_{eu} is above this value

- The number of vertical layers in each cell, as based on the euphotic depth and the topography (i.e., the maximal depth Z_{max} in the cell), is allocated as the following:

- if $Z_{max} \in \left[0, \bar{Z}_{eu} + 2sd \right]$ there is only one layer

- if $Z_{max} \in \left[\bar{Z}_{eu} + 2sd, Y \times (\bar{Z}_{eu} + 2sd) \right)$ there are two layers

- if $Z_{max} > Y \times (\bar{Z}_{eu} + 2sd)$ there are three layers

with $\bar{Z}_{eu}$ the mean euphotic depth and sd its standard deviation, and Y the number of time Z_{eu} is used to define the boundary between the 2nd and 3rd layers.

2.1.4   Configuring the input forcing data set

The next step is to prepare all forcing data in appropriate formats. As it is seen from the Fig. 3.1, running the model requires multiple input data such as physical and biogeochemical forcing, biomass of mid-trophic groups and fishing data (i.e. catch, effort and eventually size frequency of catch). Most of the files used by SEAPODYM are binary files of the format .dym (see Appendix), which can be viewed with another software with GUI called SeapodymView. Notice that the next release of SEAPODYM application and SeapodymView will work with oceanographic data stored in NetCDF files. Files containing sparse matrices (such as with fishing data) are written in ASCII format (see below).

Physical and biogeochemical data

The environmental data set to run a simulation includes temperature (^{o}C), u and v components of currents ($m\ s^{-1}$) for physical fields and primary productivity (mmol C or N $m^{-2}\ day^{-1}$), dissolved oxygen concentration ($ml\ l^{-1}$) and euphotic depth for biogeochemical fields. Primary production is vertically integrated while temperature and currents are averaged in the three layers using their definition related to the euphotic depth.
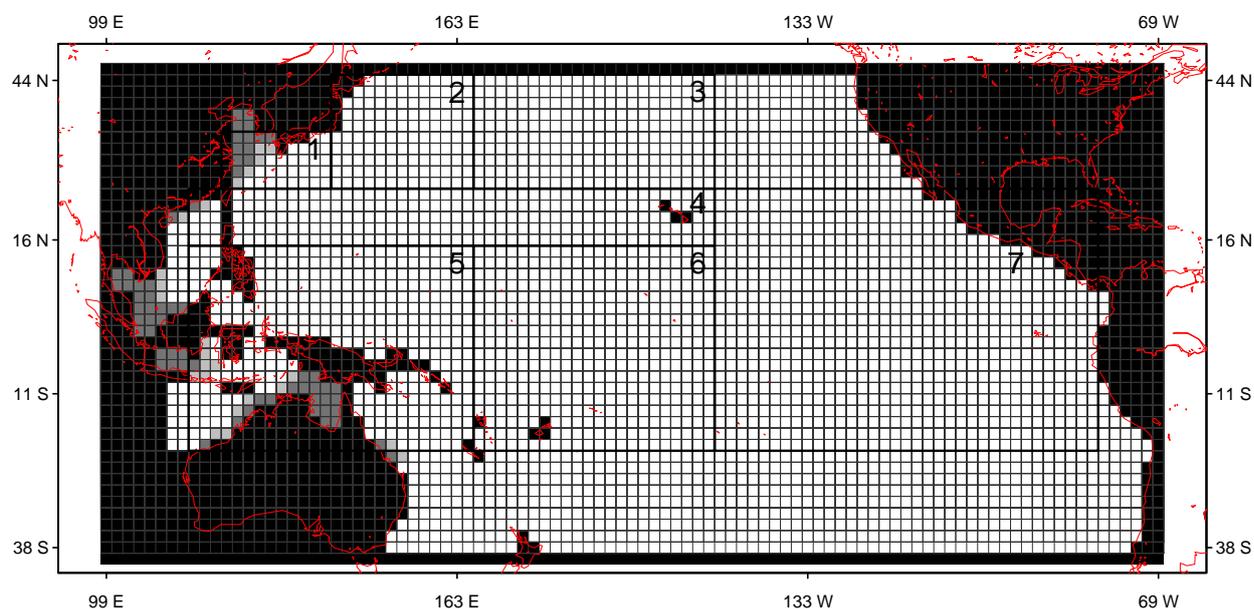
Figure 2.1: Computational domain with 3-layer mask and 2 degree grid (black color - land, dark gray - one vertical layer, light gray - two vertical layers, white - three vertical layers.

For three vertical layers, there are nine files with physical oceanographic data, three files with dissolved oxygen data and one for the primary production. Each file contains the mask and domain information, which should correspond to the model configuration. In addition, monthly climatological files should be computed from available time series to be used during the spin-up and forecast phases. The paths to all these files, together with the mask and output directory, are provided in the section 'Model INPUT-OUPUT' in the parameter file (see Appendix C).

Mid-trophic level potential biomass

The modeling of mid-trophic groups (i.e., micronekton) is done separately with the same input data set (temperature, currents and primary production). A simulation of mid-trophic levels produces six files with predicted biomass (Fbiom_*.dym) in $g\ m^{-2}$ for epipelagic, mesopelagic, bathypelagic, migrant-mesopelagic, migrant-bathypelagic and highly migrant-bathypelagic groups. Climatological files are also necessary for spin-up simulation phase. Notice that the following naming must be obeyed.

```
/forage/
./Fbiom_epi.dym
./Fbiom_meso.dym
./Fbiom_mmeso.dym
```

```
./Fbiom_bathy.dym
./Fbiom_mbathy.dym
./Fbiom_hmbathy.dym
./mclFbiom_epi.dym/
./mclFbiom_meso.dym
./mclFbiom_mmeso.dym
./mclFbiom_bathy.dym
./mclFbiom_mbathy.dym
./mclFbiom_hmbathy.dym
```

The path to the folder should be provided in the 'INPUT-OUTPUT' section of the parameter file (tag `strdir_forage`). Parameter file also includes a section with the parameters used to run the mid-trophic levels simulation called 'FORAGE MODEL PARAMETERS', however those parameters will not be used during optimization procedure or uncoupled simulation of tuna population dynamics.

Fishing data

All fishing data files should be in a folder given by `strdir_fisheries` tag in the parfile. Besides, the names of the file with original fisheries data and effort data redistributed to the model resolution (to compute the fishing mortality) must be given explicitly. The list of files is provided in the parameter file (cf Appendix C) in the section 'FISHERY DATA AND PARAMETERS'.

Catch and effort.   While the fishing effort needs to be applied at the spatial resolution of the grid to compute fishing mortality, the likelihood function uses predicted and observed catch at the resolution of the original data. In an ideal case where resolution of data and model is the same, the program only needs raw fishing data file with all operating fisheries and their effort E and catch C for single or multiple, N species with date and geographic position. The file in ASCII format has the following structure:

```
Nb of fisheries
Nb of records for fishery f1 Nb_recs f2 ...
fishery year month day gear latC lonC resolution E C_1 ... C_N
```

where `latC` and `lonC` are latitude and longitude of the center of a region and resolution is coded as follows:

code	0	1	4	5	F	T
resolution	10x10	1x1	4x10	5x5	5x10	10x20

Since fishing data are often given at different spatial resolutions and we need to apply fishing

effort at the model resolution to account for fishing mortality correctly, it is better to prepare fishing data files before simulation using GMB, which will take into account the grid and land mask information within the data redistribution procedure. Output fishing data files produced in GMB have binary format, with extension `.bin`. Note that if the model resolution coincides with the fishing data resolution there is no need to prepare .bin files, a single ASCII file with all fisheries data will be enough.

Size frequencies. It is strongly recommended to use size frequencies data by fishery for parameter estimation, however optimization Data for different fisheries are usually sampled quarterly, however the spatial resolution can vary with time and fishery. In the header of LF file for SEAPODYM we list all unique regions which were used in data sampling. The file format is the following:

```
Nb of regions     Nb of fisheries     Nb of records
1 lonE lonW latS latN                                    coordinates of region 1
2 ...                                                    coordinates of region 2
...
Number of size bins    Minimal size in the sample    bin length
year quarter month fishery region LF[1] .. LF[Nbins]
```

2.1.5   Parameter set-up

See the example parameter file with setup of parameters for South Pacific albacore tuna population in Appendix.

Simulation parameters The first section is devoted to the parameters that specify dimensions of the model variables (parameters **latitudeMin**, **latitudeMax**, **longitudeMin**, **longitudeMax** for domain size and **save_first_yr**, **save_last_yr** for the time period of simulation) and resolution of the model discretization, i.e. grid (**deltaX**, **deltaY**, in naut. mi. units) and time step (**deltaT**, in days)). Parameter **iterationNumber** gives pseudo time step for ADI numerical solver. If the flag **tuna_spinup** is set to '0' then spinup will be skipped, but in this case initial distributions should be provided in the **strdir_init** directory (see below).

deltaT *value*=30   resolution in days of computational time step

Be careful specifying all paths to the data directories correctly and verify that all necessary files are provided. The main run directory is specified in **strdir**.

In summary, before starting a simulation you must prepare:

1. Land mask, where the land is marked as '0' and ocean layers elsewhere (see above). The mask can be produced with GMB software (jar file and etopo2 maps are provided as utility for SEAPODYM).

2. Eventually, a topographic index can be used with current version to add a topographical component in the habitat definition, e.g., to restrict movements of adult cohorts to the areas with very shallow topography. Provide the path to the file in the parameter file (cf appendix). This file can also be computed for the model grid using GMB.

3. Prey fields, i.e., the predicted mid-trophic potential biomass of the 6 functional groups that are taken into account as an external forcing. The program will search precomputed with seapodym_forage sub-model (executable is provided within the package) forage distributions in the directory specified in **strdir_forage**. If this field is empty or even not existing then the files should be in the main run directory **strdir**. There must be six files (epi-, meso-, bathy-, mmeso-, mbathy- and hmbathypelagic) for the time series of the simulation and corresponding climatological files (monthly means over time period of simulation).

4. Fishing data files (see above). They are into the directory specified by **strdir_fisheries**.

5. Files with time series of modeled physical data such as currents and temperature, one file per vertical layer. The time step of the data should correspond to the time step of the model numerical scheme (**deltaT**)

6. Biogeochemical data - primary production and oxygen. Notice that if the oxygen data are available only as climatology (monthly or quarterly means) then the flag **type_oxy** should be set to 0.

7. Climatological files for all environmental data that will be used during spinup in order to generate initial conditions.

The model output will be saved in sub-directory **strdir_output**. Thus, the run directory has a simple structure: ./oceanographic_input
All model parameters in parfile have the same names as in the code. The correspondence between coded names and model notations are shown in Table 2.1. You can start working with initial guess parameters and their bounds given in the table below. In the parfile the parameter that has the tag *variable* can be estimated. If the tag < use = "true"> the parameter will be used as the control variable in optimization, otherwise, it will be static, i.e. fixed at its initial guess value. Notice that in the current version all parameters of the MTL sub-model are static.

To be sure that the parfile will be read correctly, verify that all tags have opening < and closing / > brackets.

Table 2.1: Control variables of the constrained optimization problem, imposed lower ($\underline{\theta}$) and upper ($\overline{\theta}$) boundaries and initial guess values (θ^0). Example is provided for pacific skipjack (from Senina et al., 2008.

θ	Coded name	Description	$\underline{\theta}$	$\overline{\theta}$	θ^0
$\bar{m}_p$	Mp_mean_max	maximal mortality rate due to predation,	0	1	0.5
β_p	Mp_mean_exp	slope coefficient in predation mortality,	0	0.5	0.057
$\bar{m}_s$	Ms_mean_max	maximal mortality rate due to senescence,	0	1	0.5
β_s	Ms_mean_slope	slope coefficient in senescence mortality,	-0.5	0	-0.167
A	Ms_mean_half	threshold age (in month) of tuna for senescence mortality,	20	40	31.0
ϵ	M_mean_range	variability of tuna mortality with habitat quality,	0	1	0.5
σ_0	a_sst_spawning	standard deviation in the temperature function of I_0,	2	4	3.5
$T_0^\star$	b_sst_spawning	optimal surface layer temperature for juveniles,	28.5	31.5	30
α	alpha_spawning	half saturation constant for the food to predator ratio in the spawning index,	0	5.0	0.1
σ_T	a_sst_habitat	standard deviation in temperature function of $I_{2,a}$,	1	3	2
$T_K^\star$	b_sst_habitat	optimal temperature for oldest tuna,	25	28	26.0
γ	a_oxy_habitat	slope coefficient in the function of oxygen,	-10	0	-8
$\hat{O}$	b_oxy_habitat	threshold value of dissolved oxygen,	0.1	3.0	1.0
c	c_diff_habitat	coefficient of diffusion variability with habitat index,	0	1	0.1
V_m	MSS_species	maximal sustainable speed (in body length) of tuna,	0	2	1.0
R	nb_recruitment	maximal number of larvae at large spawning biomass of adults,	0	2	0.5
b	a_adult_spawning	slope coefficient in Beverton-Holt function,	0	2	1.5
q_f	q_sp_fishery	catchability of the fishery f, Eq. 1.25	0	0.1	0.001
ς_f	s_slope (s_sp_fishery)	steepness of selectivity function, type I,	0	2.0	0.41
$\hat{l}_f$	length_threshold (s_sp_fishery)	threshold fish length,	20	70	42
μ_f	right_asymptote (s_sp_fishery)	parameter of non-symmetric Gaussian function for selectivity	0	0.5	0.002

2.2 First numerical experiment

To get some help on the command options for running SEAPODYM, use the command line:
> seapodym –help.
The information provided will be:
Usage: seapodym [options] parfile
Options:
-h, –help Print this message and exit.
-H, –hessian Compute Hessian matrix.
-p, –projection Compute 2D-projection of the likelihood on a grid specified in parfile.
-s, –simulation Run simulation without optimization.
-sa[=FLAG] Make sensitivity analysis. By default[=0] sensitivity function takes model predictions only.
–sensitivity-analysis[=FLAG] If FLAG=1 the sensitivity function takes both predictions and observations.
-t[=FLAG] Perform identical (by default, or FLAG=0) twin experiment.
–twin-experiment=[FLAG] If FLAG=1 the noise will be added to the artificial data.
-v, –version Print version number and exit.

It can be useful before to start numerical experiments with data assimilation, to test a simulation with all parameters values fixed to their best guess estimates. In that case the option -s can be used when starting the program, and all parameter values will be fixed even if their tag in the parfile are $<$ use $=$ "true"$>$.

2.3 Analyzing model outputs

A simulation produces different outputs. First, the .dym files with spatial distribution over the time series of simulation. They can be visualized and analyzed using the software SeapodymView. In the next version these files will be provided with the NetCDF format.

Biomass (in $gWWm^{-2}$) aggregated over different life stages of the species:
- sp_larve.dym // Biomass of larvae (from age 0 to one month)
- sp_juvnl.dym // Biomass of juveniles (from age 1 month to 3 month)
- sp_recru.dym // Biomass at age of recruitment
- sp_young.dym // Biomass of young (from age 2 quarter to age at first maturity)
- sp_adult.dym // Biomass of adults (from age mature to age max)
- sp_totbm.dym // Total biomass (from age 2 quarter to age max)

Total predicted and observed catch of the species:
- sp_Cobs.dym // unit ?

- sp_Cpred.dym // unit ?

Movement components of the species:
- sp_u_tot.dym // Total zonal velocity of the species (sum of current u and fish advection) in $m\,s^{-1}$ naut. mi/d ??
- sp_v_tot.dym // Total meridional velocity of the species (sum of current v and fish advection) in m, s^{-1} ?
- sp_diffusion.dym // Diffusion component of the species in m^{-2}, s^{-1} ?

The other files are in a ASCII format (.txt) and provide aggregated data over domain or regions if such regions are defined in the last section 'AGGREGATION ZONES' of the parameter file (cf appendix).

- SumDym.txt // time series of population biomass by life stage, predicted versus observed catches, fishing effort and CPUEs
- SumQArea.txt // regionally aggregated biomass (this is used to compare with MFCL outputs)
- sp_Spatial_Corr.txt // spatial correlations between predicted and observed catch and catch rates
- sp_LF_Q_sum.txt // Predicted and observed quarterly length frequencies summarized by fishery and region

If the simulation is runnning in optimization mode, the additional 'statistics.txt' ASCII file is produced. This file and all other ASCII files can be easily used with different softwares (e.g. EXCEL, R, ...) for producing plots to help in the analysis of the simulation experiment (see ex. in last chapter).

Chapter 3

Numerical experiments and data assimilation

3.1   Nonlinear function minimization

Numerical approximations of the system of equations and discrete ageing relationships define the simulation model of population dynamics. The partial derivatives of ADRs are approximated by second order finite differences with upwind differencing of advective terms (see discretization scheme in Sibert *et al.*, 1999). Boundary conditions are implemented in the discretization scheme. The complex boundary of the domain is presented by the land mask. The resulting algebraic problem is solved using the alternate direction implicit (ADI) method, with a time step of $\varDelta t = 1$.

3.2   Initialization

Initial conditions are generated by the following "spin-up" process: starting from uniform zero spatial distribution, the population density is modeled using a climatological environment (generated over the period of the time series used) during first $3\varDelta\tau_0 + K\varDelta\tau_a$ time steps. Every time step, a new larval source is computed using the temperature only (i.e., $S_{J_0} = R\Phi(T_0)$). The duration of the spin-up is set up to match the total lifespan of the species population, so that the density of each age class can be computed at the end of the climatological run. After spin-up, the simulation continues with actual forcing fields for another 2 years in order to reduce the influence of initial climatological forcing. Final distributions are saved for later use as initial conditions for optimization experiments. As the initial distributions play an important role in parameter estimation process, we then repeated the same procedure several times re-generating initial state of the model using the optimized parameters.

3.3 Fisheries Data assimilation

Data assimilation is based on fishing data from multiple fisheries. To maintain the time of computation at a reasonable level, it is necessary to find a balance between the number of fisheries, the size of the domain, the spatial and temporal resolution, and the extension of the time series used in the optimization experiment. Since each fishery is characterized in the model by one constant catchability coefficient and a selectivity function, it is also important to aggregate homogeneous fishery data. Typically, for tuna species, monthly spatially distributed data on fishing are available for effort E_{tfij} (in days or nb hooks) and catch C_{tfij}^{obs} (in weight). Size frequency data is usually aggregated by fishery and large area, e.g., 10x10 deg. squares.

As an example, the study for Pacific skipjack tuna in Senina et al. (2008) used six generalized gear types or "fisheries" defined by unique values of the "catchability coefficient", q_f with $f = 1, 2, \ldots, 6$ corresponding to four WCPO (PLSUB, PLTRO, WPSASS, WPSUNA) and two EPO fisheries (EPSASS and EPSUNA), while seasonal size composition of the catch was available for each fishery aggregated over 7 spatial regions. In addition, the computational domain was restricted to $\Omega = \{x \in (99°E, 69°W), y \in (45°N, 39°S)\}$ since this is the region where skipjack catches have been recorded during the 1950-2005 period (see Fig. **??**), and a coarse grid with $\Delta x = \Delta y = 120$ Nmi (nautical miles) was used in these first experiments.

3.3.1 Model predictions

The predicted catch, C_{tfij}^{pred}, at time t for fishery f is computed in the model using observed fishing effort E_{tfij} at location (i, j) by

$$C_{tfij}^{pred} = q_f E_{tfij} \sum_{a=1}^{K} s_{fa} w_a N_{aij} \Delta x \Delta y,$$

where w_a is the mean weight of fish in the a-th cohort.
The predicted proportion at age a in the catch at time t for fishery f in region r is

$$Q_{tfar}^{pred} = \frac{s_{fa} \sum_{i,j \in r} E_{fij} N_{aij} \Delta x \Delta y}{\sum_{a=1}^{K} s_{fa} \sum_{i,j \in r} E_{fij} N_{aij} \Delta x \Delta y}.$$

3.4 Maximal likelihood approach

SEAPODYM explicitly describes spatial dynamics of pelagic fish populations influenced not only by intrinsic population dynamics processes, but also by extrinsic environmental
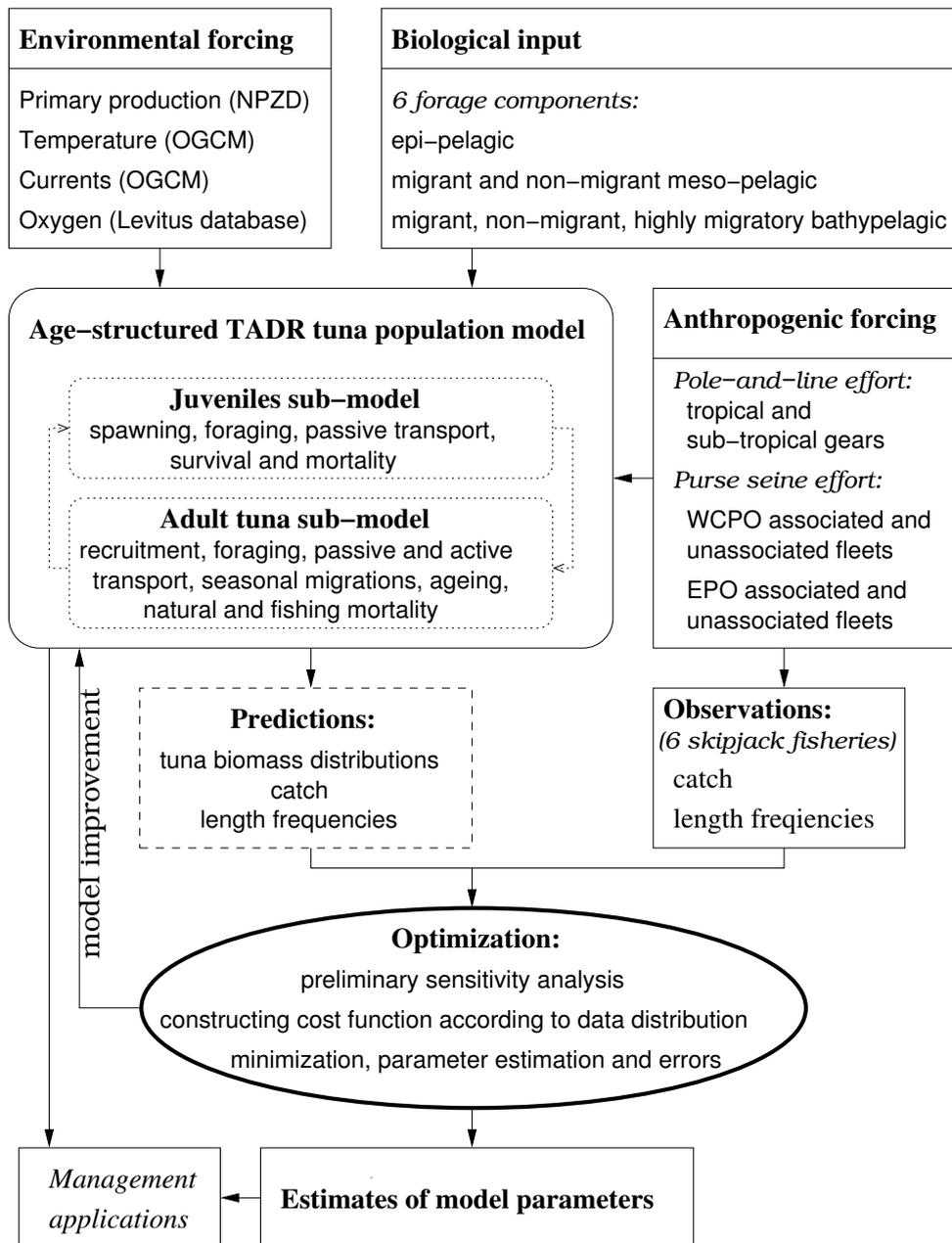
Figure 3.1: General scheme of the model with optimization approach.

variability. Predictions of the model strongly depend on environmental forcing which is taken by the model as the input. However, in order to have confidence in the model predictions, particularly the adequacy of the population dynamical responses to environmental variability, we need to combine simulations with quantitative optimization (see the lower part of Fig. 3.1) and to express how well the model describes the observational data.

3.4.1 Likelihood functions

We use the maximum likelihood method to estimate model parameters θ_k that would allow the model predictions to approach observations.

Catch likelihood components Different functions can be used to represent the distribution of the catch data. One frequently used is Poisson distribution. An example is given in the skipjack application in Senina et al. (2008) with the WCPO purse seine fleet that is targetting this species. However, when fishing data contain many zeros, as for the skipjack catch data from Eastern Pacific Ocean (EPO) purse seine fleets that target mostly yellowfin tuna, it may be more appropriate to use other distributions, e.g., a negative binomial distribution with zero inflation.
This yields the following likelihood components with Poisson distribution:

$$L_1\left(\boldsymbol{\theta}|C^{obs}\right) = \prod_{tfij} \frac{C_{tfij}^{pred\,C_{tfij}^{obs}} e^{-C_{tfij}^{pred}}}{C_{tfij}^{obs}!},\ f = 1,2,3,4; \tag{3.1}$$

and for negative binomial distribution with zero inflation:

$$L_2\left(\boldsymbol{\theta}|C^{obs}\right) = \begin{cases} \displaystyle\prod_{tfij}\left(p_f + (1-p_f)\left(\frac{\beta_f}{1+\beta_f}\right)^{\frac{\beta_f C_{tfij}^{pred}}{1-p_f}}\right),\text{ if } C_{tfij}^{obs} = 0, \\ \displaystyle\prod_{tfij}\left((1-p_f)\frac{\Gamma\left(C_{tfij}^{obs}+\frac{\beta_f C_{tfij}^{pred}}{1-p_f}\right)}{\Gamma\left(\frac{\beta_f C_{tfij}^{pred}}{1-p_f}\right)C_{tfij}^{obs}!}\left(\frac{\beta_f}{1+\beta_f}\right)^{\frac{\beta_f C_{tfij}^{pred}}{1-p_f}}\left(\frac{1}{1+\beta_f}\right)^{C_{tfij}^{obs}}\right),\text{ if } C_{tfij}^{obs} > 0, \\ \hfill f = 5,6. \end{cases} \tag{3.2}$$

where the parameters β_f and p_f, are the negative binomial parameters (showing how much variance exceeds expected value) and probability of getting a null observation, respectively. Both are estimated in the optimization process.

Length frequencies likelihood components. We assume that fish lengths at catch are normally distributed, which gives the following contribution from length frequency data to the negative log-likelihood:

$$-L_3\left(\boldsymbol{\theta}|Q^{obs}\right) = \frac{1}{2\sigma_Q^2}\sum_{t=1}^{T}\sum_{f=1}^{6}\sum_{a=1}^{K}\sum_{r=1}^{7}\left(Q^{pred}_{tfar} - Q^{obs}_{tfar}\right)^2, \tag{3.3}$$

where the proportion at age a in the catch computed from the observed length-frequency data is $Q^{obs}_{tfar} = \frac{N^{obs}_{tflr}}{\sum_l Ntflr}$, $l \in \left[\underline{l_a}, \overline{l_a}\right)$. N^{obs}_{tflr} is the number of fish of length l in region r that belongs to the cohort of age a. The variance σ_Q^2 is assumed to be a constant value (1.5 cm). The negative log-likelihood function, $L^- = -\ln(L)$, to be minimized is thus the sum of the three components described above, i.e.,

$$L^- = -\ln L_1\left(\boldsymbol{\theta}|C^{obs}\right) - \ln L_2\left(\boldsymbol{\theta}|C^{obs}\right) - L_3\left(\boldsymbol{\theta}|Q^{obs}\right). \tag{3.4}$$

Instead of setting penalties to the boundaries of $\boldsymbol{\theta}$, we chose to perform a constrained minimization through parameter scaling (see e.g., Bard, 1974; Vallino, 2000). The latter implies that the optimization routine operates in the unbounded parametric space that is mapped to the bounded one with the transformation $\theta_k = \underline{\theta_k} + \left(\bar{\theta}_k - \underline{\theta_k}\right)\left(1 + sin\frac{\pi\theta'_k}{2}\right)$, i.e., variable θ'_k can vary from $-\infty$ to ∞ while θ_k remains within the imposed bounds.

3.5 Numerical experiments

Despite the minimum number of parameters used in the model to describe the full spatial population dynamics of a species, the model will likely have difficulties to converge quickly toward the best solution. Preliminary optimization experiments can reveal different problems, especially if there are bad or too sparse fishing data.

First, it is recommended to reduce the influence of initial conditions (spatial distributions of cohorts produced by the model with initial guess parameters) on the results of minimization procedure. this can be done by skipping the first predictions in the likelihood computation. For the skipjack application (Inna et al. 2008), the first 6-month predictions in each optimization experiment were excluded from the likelihood function (parameter nb_step_to_skip value="6"). Three optimization experiments were conducted also, that spanned three different time periods: 1980-1990 (E1), 1990-2005 (E2), and the total 25 year range from 1980 to 2005 (E3) to determine the extent to which the estimated parameters depend on data from different periods and different initial conditions, and consequently, how model predictions change as a result of assimilating data over different time periods.

Environmental forcing variables have a strong impact on the results and can produce biases if they are not realistic enough. Continuing with our skipjack example, we noticed that the coarse spatial resolution (2 deg) of the grid being used and maybe some other problems (e.g., lack of ice model) produced a bias in the predicted seasonal peak of temperature in

the forcing field for the Kuroshio region. This temperature variability is however of primordial importance to determine seasonal population migrations in this area (Nihira, 1996). Therefore with this configuration SEAPODYM was not adequately describing the seasonal skipjack migration through the Kuroshio extension and attempts to assimilate data within this geographic area would bias parameters of habitat indices. Consequently, the sub-tropical pole-and-line fleet catch data were excluded from the likelihood calculation (but not from the computation of fishing mortality) and the catchability coefficient (fishing mortality) for this fleet was fixed.

Correlation between different parameters is a classical issue. For example, the simultaneous estimation of the parameters of the larval recruitment on one side and the natural and fishing mortality functions on the other side can lead to biased estimates of the total population size. Namely, if we try to estimate simultaneously parameters R, b, β_P and β_S in the skipjack application, the minimization procedure tends to increase the total stock and also the computational time (i.e., the number of iterations) increases by a factor of more than three compared to experiments in which R is fixed. Such drastic increase of the number of iterations due to the release of one more parameter is an indication of strong correlation with other parameters. Considering the importance of the estimation the mortality rate for each cohort given its functional form over ages, and taking into account availability of more information for the adult cohorts in the fishing data than for larvae stage, we have chosen to fix the recruitment parameters. But we reduced also the number of mortality function parameters by fixing the coefficient of variability with habitat index ε at its guessed value 0.5, because sensitivity analysis (see below) indicated that the model configuration was barely sensitive to this parameter.

3.6   Sensitivity analyses

Sensitivity analyses are a useful approach to reveal which parameters can be estimated from available data and which can not. If model predictions are insensitive to some parameters, it is unlikely that they will be determined uniquely from available observations and should, therefore, be removed from the optimization. Two types of sensitivity analyses can be performed with this version of SEAPODYM.

Sensitivity to parameters.   The first type of analysis examines how the predictions of the model are sensitive to its parameters. For this purpose we simply need to construct a function of the model solution, which represents model predictions (see, e.g., Worley, 1991). Then, the measures of sensitivity can be computed using precise gradients obtained from adjoint calculations. Since two types of data are assimilated within the model, i.e., catch and

length-frequencies, we construct the following functions:

$$R_1 = \sum_{tfij} \left(C_{tfij}^{pred} \right)^2, \ \ R_2 = \sum_{tfar} \left(Q_{tfar}^{pred} \right)^2, \ \ \text{r=1,...,7.} \tag{3.5}$$

Then we define two measures of relative sensitivity $\xi_1(\theta_k^0)$ and $\xi_2(\theta_k^0)$ for corresponding model predictions and each initial guess parameter θ_k^0 as follows:

$$\xi_1(\theta_k^0) = \frac{1}{R_1} \frac{\partial R_1}{\partial \theta_i^0}, \ \ \xi_2(\theta_k^0) = \frac{1}{R_2} \frac{\partial R_2}{\partial \theta_i^0}. \tag{3.6}$$

▷ To run this type of analysis, use the option -sa[0] in the command line to execute SEAPODYM.

Sensitivity to objective function. The second type of sensitivity analysis examines whether the objective function (which incorporates both predicted and observed data) is sensitive to model parameters. We compare values of likelihood at some found minimum $\boldsymbol{\theta}^\dagger$ to those evaluated at boundaries of parameter space (Vallino, 2000). We define two further measures of relative sensitivity:

$$\xi_3(\theta_k^\dagger) = \frac{L^-(\boldsymbol{\theta}^\dagger + \delta\bar{\theta}_k \cdot \mathbf{e}_k) - L^-(\boldsymbol{\theta}^\dagger)}{L^-(\boldsymbol{\theta}^\dagger)}, \ \ \xi_4(\theta_k^\dagger) = \frac{L^-(\boldsymbol{\theta}^\dagger - \delta\underline{\theta}_k \cdot \mathbf{e}_k) - L^-(\boldsymbol{\theta}^\dagger)}{L^-(\boldsymbol{\theta}^\dagger)} \tag{3.7}$$

where $\delta\bar{\theta}_k = \bar{\theta}_k - \theta_k^\dagger$, $\delta\underline{\theta}_k = \theta_k^\dagger - \underline{\theta}_k$ and $\mathbf{e}_k$ is a standard basis vector with 1 in the k-th element and 0 elsewhere.

▷ To run this type of analysis, use the option -sa[1] in the command line to execute SEAPODYM.

For both -sa[ ] options, outputs are directly shown on the terminal screen at the end of the simulation and can be recorded in a text file if needed to produce a table (Table 3.1) or a plot (e.g., Fig 3.2). It is recommended to run multiple sensitity tests from experiments with different initial conditions, time series of simulation, and estimated values of parameters. This approach will help in detecting persistent patterns, e.g., parameters that show consistenly low or conversely really high sensitivity, and hopefully the reason of such patterns. If the problem cannot be solved, e.g., by increasing the period of simulation to include new data or testing a better environmental forcing data set, then the parameter(s) with low sensitivity should be fixed to their best guess-estimate and removed from the optimization.

Sensitivity analyses only give a tentative representation of observable and non-observable parameters unless the entire parameter space is explored thoroughly. practically, performing exhaustive sensitivity analysis, i.e., exploring entire likelihood hyper-surface in n-dimensional parametric space is impossible. Consequently, some non-estimable parameters are likely unrecognized, but large uncertainties obtained from the error analysis (see below) are good

Table 3.1: Example of outputs from the sensitivity analyses.

N	parameter	relative sensitivity
1	Mp mean exp(0)	8.72181
2	Ms mean max(0)	-1.11762
3	a sst spawning(0)	2.54945 e-10
4	b sst spawning(0)	0.00656738
5	a sst habitat(0)	3.39733e-18
6	b oxy habitat(0)	-1.0208
7	MSS species(0)	-0.070417
8	nb recruitment(0)	0.0047095
9	...	
10	...	

Sensitivity measures for model parameters
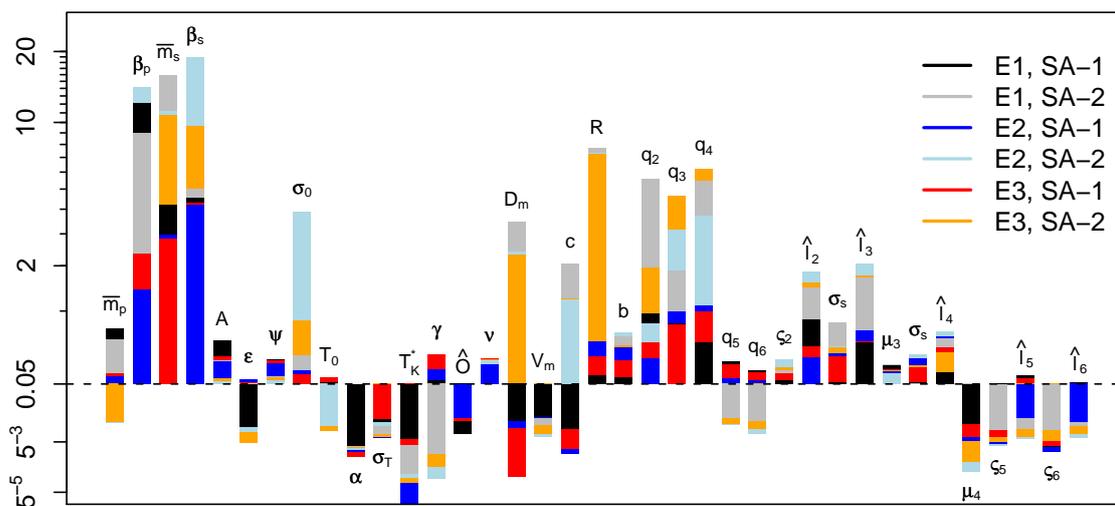
Figure 3.2: Log-scaled measures of sensitivity obtained for estimated parameters from three different experiments E1, E2 and E3. The values below dashed line correspond to less then 5% sensitivity of either predictions (SA-1, i.e., $max|\xi_1, \xi_2|$) or objective function (SA-2, $max|\xi_3, \xi_4|$) to corresponding parameter. Redrawn from Senina et al. (2008).

indicators of poorly determined parameters. Unfortunately, one cannot prove that minima found in these experiments are not local minima. It is possible, however, to evaluate the reliability of the solutions by conducting so-called "identical twin experiments".

3.7 Identical twin experiments

Demonstrating that a solution found by numerical optimization is a global solution and not a local minimum is a difficult and well-known problem in all non-linear optimization problems and data assimilation. In order to verify that both the model and the method allow to estimate chosen parameters using the available amount of observations, it is possible to conduct "identical twin experiments". These tests consist in estimating parameters from artificial data series constructed from predictions given by the model. If optimization works well with the model and experiment set-up, then after sufficient perturbation of optimal parameters we should be able to retrieve them, because they determine known a-priori solution represented in the artificial data series. Twin experiments conducted for the skipjack application have shown that the control parameters were successfully recovered with small relative errors $\varepsilon < 0.001$ due to the computer round-off error. The evolution of the parameters during minimization process can be plotted like in the figure 3.3 to illustrate how quickly the parameters were recovered, and which parameters are the most difficult to be correctly estimated.

▷ To run a twin-experiment:
- create an artificial fishing data set from a given model configuration, with or without adding noise to the data, and use it as a new fishing data set.
- change the initial values of the parameters to be estimated in the parameter file.
- use the option -t[0] in the command line to execute SEAPODYM.
- use the statistics.out output file to check the result and produce plots (e.g., Fig. 3.3).

3.8 2D projection of likelihood function

Sometimes it is useful to explore the likelihood function visually, looking at different projections over the parameter space. It gives the clue what the sensitivity of a certain parameter is, whether the boundaries of the parameter should be modified, how the imposed penalty functions change the shape of the likelihood etc. Plotting the cost function over a couple of variable parameters is also a simple way to visualize the localization of found solution. For example, the likelihood function can have multiple minima and the likelihood projection plot can reveal the local minimum problem and help to navigate the minimization experiments further. The numerical instability problems, or the high-mode nonlinearity of model solutions, and/or noisy observations can result in abrupt, spiky surfaces of the cost function and
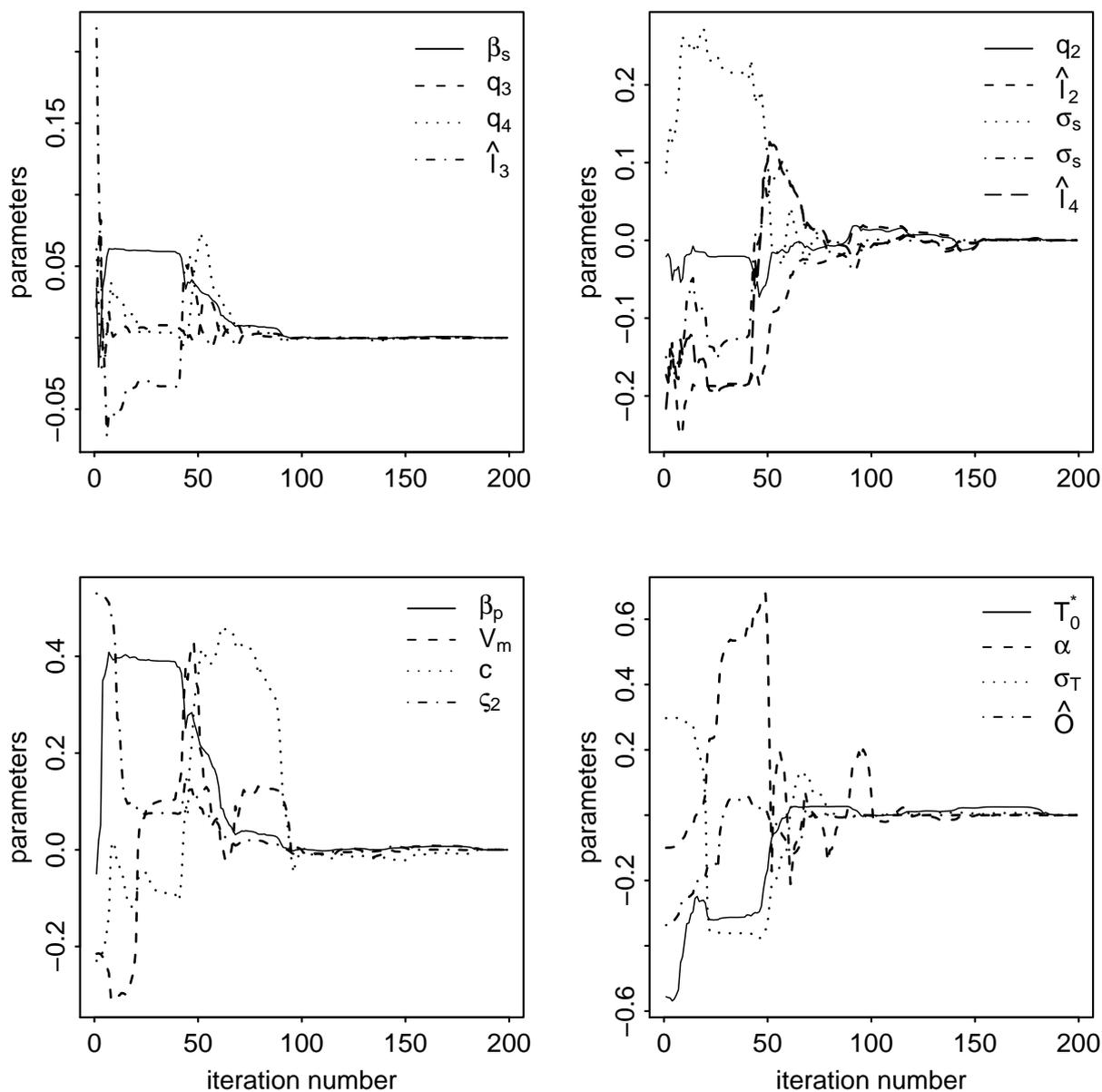
Figure 3.3: Evolution of control parameters during twin data experiment conducted for the artificial data simulated with E2 parameter set. Parameters are grouped by their sensitivities in descending order, except β_p for which high sensitivity was calculated by both SA-1 and SA-2 analysis.

hence impede the convergence of the minimization method. Figure 3.4 illustrates three examples of the the cost function being (a) ideal case of quadratic shape, (b) two local minima and (c) extremely non-linear surface due to high diffusion rates instability problem.

All you have to do for computing and further plotting the 2D projections of likelihood function is to add the following lines to the parfile:

```
<hyperspace_projection>
    <variables nb="2"/>
    <var1 name="alpha_spawning" nsteps="5" />
    <var2 name="nb_recruitment" nsteps="5" />
</hyperspace_projection>
```

and run the program with the command

```
seapodym -p parfile.xml
```

In the example above the program will split the parameter plane bounded by values specified by **min/max** tags into 25 grid cells and run simulations to evaluate the likelihood with gridded parameters.

3.9   Parameter estimates, variance and errors

Unfortunately, there is no simple way to evaluate the uniqueness of the estimated parameters in nonlinear problems especially for ecosystem models due to their extreme complexity, high dimension of the objective functional and scarcity of available observations (Robinson and Lermusiaux, 2002, Vallino, 2000). Different experiments using different time periods and data sets will produce differences in the parameter estimates. The approach of perturbing the model parameters and restarting the experiment allows to check that the model has converged towards the minimum of the experiment, but it is not possible to assert that found solutions are not local minima. However considering the dimension of the minimization problem and the computer time to perform one experiment, it is also unrealistic to envisage an exhaustive study that would allow us to conclude that computed solutions are global.

We can, however, determine whether the estimated parameters were well determined at the minima detected by minimization routine. For this, wee need to compute the variance of the estimated parameters from the inverse of the Hessian matrix, i.e., $C = \mathbf{H}^{-1}$ (Bard, 1974), where $\mathbf{H} = \frac{\partial^2 L^-}{\partial \theta_i \partial \theta_j}$, $i, j = 1, 2, ..., n$ is the Hessian matrix evaluated at the minimum of the negative log likelihood function. The diagonal elements of C provide estimates of the variance of the optimal parameters. The Hessian matrix is approximated with central finite difference using first derivatives exactly evaluated by adjoint calculations. From the variance, standard
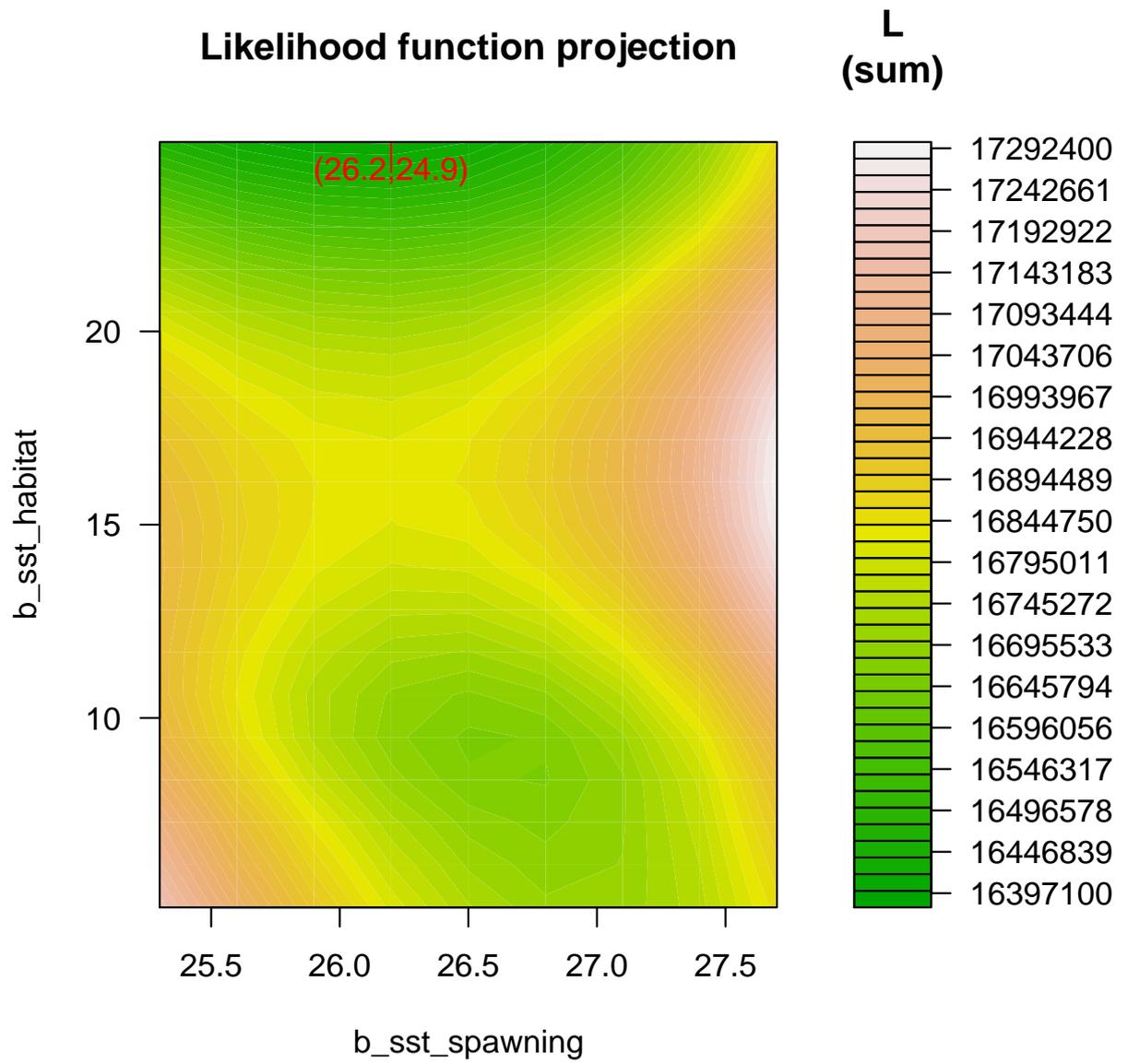
Figure 3.4: Likelihood projection examples

deviation can be calculated for estimated parameters, allowing to check the amount of uncertainty on these etimates.

Correlation coefficients between pairs of estimated parameters can also be calculated from the error-covariance matrix. This is useful to identify problems in the optimization approach and eventually to define new experiments trying to estimate the correlated parameteres separately.

▷ To compute the Hessian matrix:
- use the option -H in the command line to execute SEAPODYM.
- use the statistics.out output file to check the results and produce plots or tables.

Chapter 4

Visualization of model predictions.

Spatial distributions. SeapodymView software   For rapid access when writing or reading data during SEAPODYM simulations, a special binary format (using IEEE float) was defined and is identified by the extension ¡.dym¿. A DYM file is composed with a header and the series of 2-dimensional matrices representing the spatial distribution of the variable in a chronological order, each matrix being read from north to south and east to west. SeapodymView is a visualization software (developed in JAVA) allowing to display .dym format data with related options and information, and proposes several overlay options (Fig. 4.1): Circles and velocity Components based on dym files; polygons or E.E.Z. boundaries, individual tracks and fishing data (ASCII files). There are also several tools to manipulate the data, produce simple computation and export data and images. SeapodymView is provided with a separate documentation. Note that a new version of the software is under development that will be based on NetCDF format.

Time series. R routines   During simulation run the program computes aggregated biomass and catches and stores predicted time series (together with observed ones) into ASCII files (see list of files and their contents in 3.3). To plot these time series there provided the set of R routines.

It is important and very often helpful to compare the predictions made by different models, especially if they utilize conceptually different approaches. One such model which is used to estimate tuna stock for management applications is Multifan-CL, age-structured regional-based statistical stock assessment model which uses the fisheries and tag recaptures data to estimate the parameters and fit the model trajectories.

In order to write the model output to be compared with MFCL regional predictions, one need to specify the same regional structure in AGGREGATION ZONES section of the parameter file. In Appendix C see an example of regions for albacore analysis. Since for some species the only available length frequencies data are the ones being used in MFCL analysis, the earlier version of the code used these regions for reading and computing length frequencies. With the new fine resolution dataset the regional definition through parameter
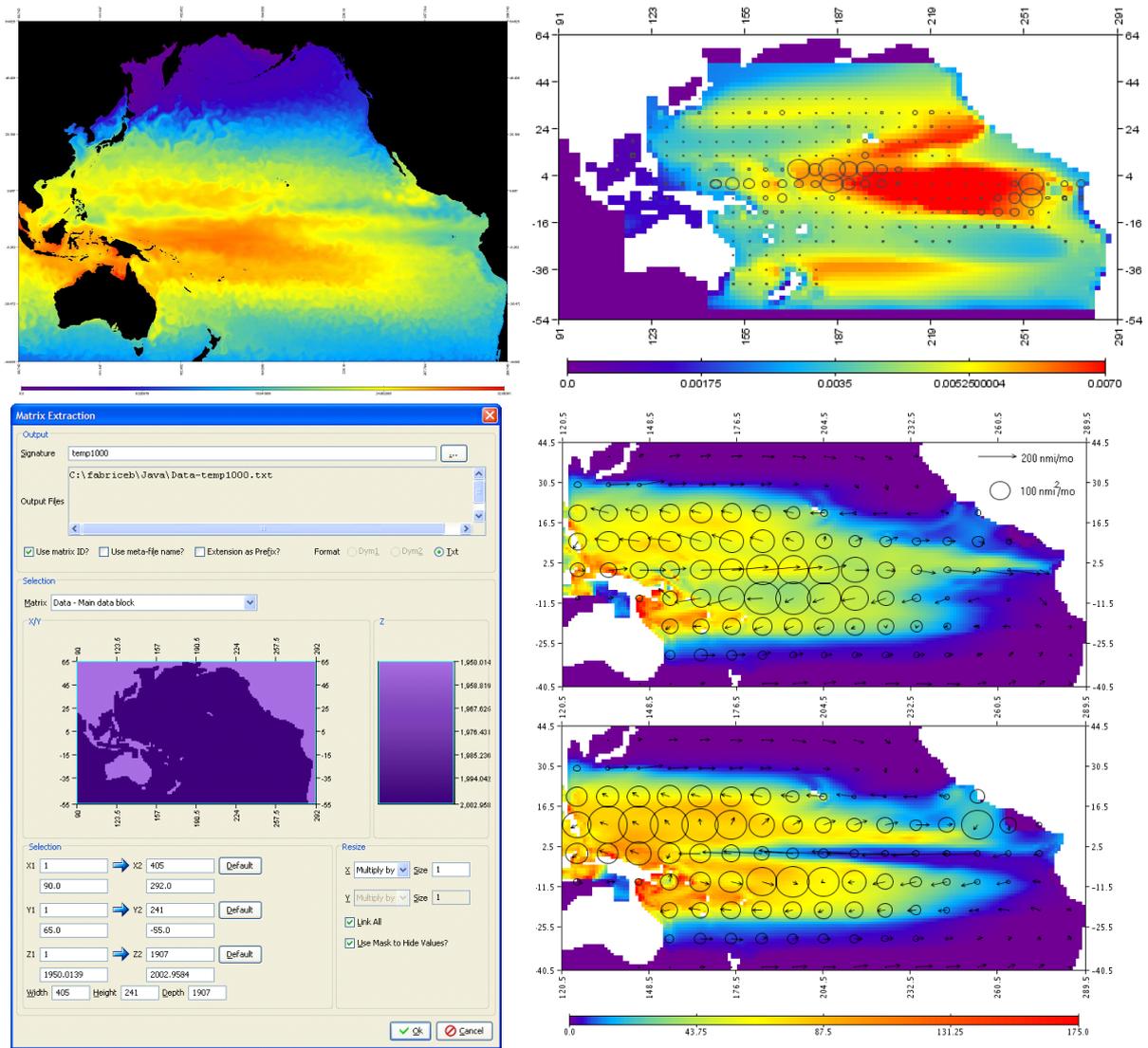
Figure 4.1: SeapodymView visualization software

file became redundant, however the code still uses the same variables and data containers, so if one reads the fine resolution size data from file *file_frq_data* then, for example, the variable *nb_region* will be reassigned to the number of unique regions in the data. Hence, for Seapodym and MFCL outputs comparison one need do the following:

- set tag *frq_likelihood value = 0*

- set tag *flex_regstruc value = 0*

- specify number of regions and spatial stratification in accordance with MFCL definitions

Appendix A

Model variables and parameters

A.1   Selected notations

N	Symbol	Description	Units
1	Ω, x, y	2d model domain with complex boundary and its coordinates	*degrees*
2	z	vertical layers (1) 0-100m, (2) 100-400m and (3) 400-1000m	*m*
		Environmental data	
3	$\mathbf{v}_z$	vector (u, v) of horizontal currents, averaged through each vertical layer (GCM modeled data)	*nmi/mo*
4	T_z	temperature, averaged through each layer z, (GCM data)	°C
5	O_z	concentration of dissolved oxygen, averaged through each vertical layer (Levitus database)	*ml/l*
6	P	primary production, averaged through 0-400m depth (obtained from GCM-NPZD coupled model)	mmol$Cm^{-2}mo^{-1}$
		ADR coupled model variables	
7	F_n	density of n-th forage component (food for tunas)	g/m^2
8	J_k	density of juvenile age class $k = 0, 1, 2$ of tuna population	g/m^2
9	N_a	density of adult age class $a = 1, .., K$ of tuna population	g/m^2
		Environmental(habitat) indices	
10	$\Theta_{a,n}$	accessibility of tuna cohort a to n-th forage vertical habitat	
11	I_0	spawning or larvae's habitat index	
12	I_1	juvenile's habitat index	
13	$I_{2,a}$	adult's (feeding, movement and seasonal migrations) habitat index	
		Advection-diffusion-reaction parameters	
14	V_a	vector of velocity of each tuna cohort density	*nmi/mo*
15	D_a	diffusion coefficient for each tuna cohort	nmi^2/mo
16	m_S	tuna senescence mortality	mo^{-1}
17	m_P	tuna predation mortality	mo^{-1}
18	m_F	tuna fishing mortality	mo^{-1}
19	$s_{f,a}$	selectivity functions for fishery f and age of tuna a	
		Optimization variables	
20	C_f	total monthly tuna catch by fishery f	10^3 tones
21	Q_f, r	proportion of length frequencies for fishery f and region r	
22	L^-	total negative likelihood function	

A.2 SEAPODYM parfile

parfile.xml

```xml
<?xml version="1.0"?>
<!-- check http://www.w3.org/TR/2004/REC-xml-20040204 for current xml specifications -->
<par version="1.0">

<!--==================== -->
<!--SIMULATION PARAMETERS-->
<!--==================== -->
<latitudeMin value="-54.0"/>
<latitudeMax value="4.0"/>
<longitudeMin value="130.5"/>
<longitudeMax value="288.5"/>
<nb_layer value="3"/>
<deltaX value="120"/>
<deltaY value="120"/>
<deltaT value="30"/>
<iterationNumber value="8"/>
<tuna_spinup value="0"/>
<save_first_yr value="1980"/>
<save_last_yr value="2000"/>
<nb_yr_forecast value="0"/>
<nb_step_to_skip value="12"/>

<!--==================== -->
<!--LIKELIHOOD PARAMETERS-->
<!--==================== -->
<!--type of data in the likelihood: 0 - cpue, 1 - catch -->
<data_type value="1"/>

<total_likelihood value=""/>

<!--likelihood types: 1 - concentrated, 2 - log-normal, 3 - poisson, 4 - negative binomial, 5 - zero-inflated negbin, 6 - truncated poisson -->
<!--can be different for each fishery (see below the fishery codes) -->
<likelihood_types>
    <alb> 3 </alb>
</likelihood_types>

<nb_fishery value="1"/>
<list_fishery_name>L10</list_fishery_name>

<!--Mask for fishery data usage: 0 - do not take into account-->
<mask_fishery_sp>
    <alb>1</alb>
</mask_fishery_sp>

<mask_fishery_likelihood>
    <alb>1</alb>
</mask_fishery_likelihood>

<likelihood_parameters>
    <variables use=""/>
        <L10 alb=""/>
</likelihood_parameters>

<prob_zero>
    <L10 alb=""/>
</prob_zero>
```

Figure A.1: Example xml parfile

parfile.xml

```xml
<likelihood_type value="1"/>

<!--length frequencies likelihood: 0 - do not take into account, 1 - account-->
<frq_likelihood alb="0"/>

<frq_likelihood_weight alb="0.015"/>

<!--================= -->
<!--MODEL INPUT-OUTPUT-->
<!--================= -->
<!-- DataFiles -->

<!-- Working Directory -->
<strdir value="run-pfrp-ncep-sp/"/>

<str_file_mask value="Mask_2deg_3L_SP.txt"/>
<str_file_topo value="topo_2deg_SP_50m.txt"/>

<strdir_forage value="forage/"/>

<strdir_fisheries value="fisheries/"/>

<!--type of oxygen data: 0 for seasonal climatology, 1 for time series-->
<type_oxy value="1"/>

<strfile_pp value="PP_PFRP_NCEP_SP_1948_2003.dym"/>
<strfile_u layer0="U_PFRP_NCEP_SP_L1_1948_2003.dym" layer1="U_PFRP_NCEP_SP_L2_1948_2003.dym"
layer2="U_PFRP_NCEP_SP_L3_1948_2003.dym"/>
<strfile_v layer0="V_PFRP_NCEP_SP_L1_1948_2003.dym" layer1="V_PFRP_NCEP_SP_L2_1948_2003.dym"
layer2="V_PFRP_NCEP_SP_L3_1948_2003.dym"/>
<strfile_t layer0="T_PFRP_NCEP_SP_L1_1948_2003.dym" layer1="T_PFRP_NCEP_SP_L2_1948_2003.dym"
layer2="T_PFRP_NCEP_SP_L3_1948_2003.dym"/>
<strfile_oxy layer0="O_PFRP_NCEP_SP_L1_1948_2003.dym" layer1=
"O_PFRP_NCEP_SP_L2_1948_2003.dym" layer2="O_PFRP_NCEP_SP_L3_1948_2003.dym"/>

<strfile_pp_clm value="mclPP_PFRP_NCEP_SP_1980.dym"/>
<strfile_u_clm layer0="mclU_PFRP_NCEP_SP_L1_1980.dym" layer1="mclU_PFRP_NCEP_SP_L2_1980.dym"
layer2="mclU_PFRP_NCEP_SP_L3_1980.dym"/>
<strfile_v_clm layer0="mclV_PFRP_NCEP_SP_L1_1980.dym" layer1="mclV_PFRP_NCEP_SP_L2_1980.dym"
layer2="mclV_PFRP_NCEP_SP_L3_1980.dym"/>
<strfile_t_clm layer0="mclT_PFRP_NCEP_SP_L1_1980.dym" layer1="mclT_PFRP_NCEP_SP_L2_1980.dym"
layer2="mclT_PFRP_NCEP_SP_L3_1980.dym"/>
<strfile_oxy_clm layer0="mclO_PFRP_NCEP_SP_L1_1980.dym" layer1=
"mclO_PFRP_NCEP_SP_L2_1980.dym" layer2="mclO_PFRP_NCEP_SP_L3_1980.dym"/>

<!-- Output Directory -->
<strdir_output value="output/simout/"/>

<!--===================== -->
<!--FORAGE MODEL PARAMETERS-->
<!--===================== -->
<Tr_max value="450"/>
<Tr_exp value="-0.1"/>
<inv_lambda_max value="1825"/>
<inv_lambda_curv value="-0.1"/>
<E value="0.00316"/>
<c_pp value="0.0948"/>
```

Figure A.2: Example xml parfile (continued)

```
parfile.xml
<nb_forage value="6"/>

<frg_name>epi meso mmeso bathy mbathy hmbathy</frg_name>

<source_frg epi="0.17" meso="0.1" mmeso="0.22" bathy="0.18" mbathy="0.13" hmbathy="0.2"/>
<day_layer epi="0" meso="1" mmeso="1" bathy="2" mbathy="2" hmbathy="2"/>
<night_layer epi="0" meso="1" mmeso="0" bathy="2" mbathy="1" hmbathy="0"/>

<tstep_forage value="30"/>
<sigma_fcte value="250"/>
<max_NPP value="200"/>
<b_F_habitat value="20000"/>

<!--======================== -->
<!--PREDATOR MODEL PARAMETERS-->
<!--======================== -->
<nb_species value="1"/>

<sp_name>alb</sp_name>

<season_peak alb="1"/>

<gd_threshold alb="0.035"/>

<vertical_movement alb="1"/>

<thermal_inertia alb="1"/>

<penalty_mortality alb="0"/>

<a_sst_spawning alb="1.4">
    <variable min="1.0" max="4.0" use="true"/>
</a_sst_spawning>

<b_sst_spawning alb="23.1">
    <variable min="21.0" max="30.0" use="true"/>
</b_sst_spawning>

<alpha_spawning alb="1.0">
    <variable min="0" max="10.0" use="true"/>
</alpha_spawning>

<nb_recruitment alb="0.003">
    <variable min="0.0" max="1.0" use="true"/>
</nb_recruitment>

<a_adults_spawning alb="0.8">
    <variable min="0.1" max="5" use="true"/>
</a_adults_spawning>

<a_sst_habitat alb="2.0">
    <variable min="1.0" max="3.5" use="false"/>
</a_sst_habitat>

<b_sst_habitat alb="15.0">
    <variable min="10.0" max="28.0" use="false"/>
</b_sst_habitat>
```

Figure A.3: Example xml parfile (continued)

parfile.xml

```xml
<a_oxy_habitat alb="0.01">
    <variable min="1e-10" max="1" use="true"/>
</a_oxy_habitat>

<b_oxy_habitat alb="3.68">
    <variable min="2" max="6.0" use="true"/>
</b_oxy_habitat>

<hp_cannibalism alb="5.0">
    <variable min="0" max="20" use="false"/>
</hp_cannibalism>

<sigma_species alb="0.25">
    <variable min="0" max="1.05" use="false"/>
</sigma_species>

<c_diff_fish alb="0.3">
    <variable min="0" max="1.5" use="false"/>
</c_diff_fish>

<MSS_species alb="0.5">
    <variable min="0.3" max="2.5" use="false"/>
</MSS_species>

<sp_unit_age_class_jv alb="30"/>
<sp_nb_age_class_jv alb="3"/>

<juv_length>
    <alb>3.3 8.4 10.8</alb>
</juv_length>

<juv_weight>
    <alb>0.0043 0.009 0.029</alb>
</juv_weight>

<sp_unit_age_class_yn alb="180"/>
<sp_unit_age_class_ad alb="360"/>
<sp_nb_age_class_ad alb="20"/>

<age_recruit alb="0"/>
<age_mature alb="9"/>

<length>
    <alb>16.557 29.873 41.051 50.434 58.312 64.924 70.475 75.135 79.046 82.330 87.401 90.974
93.492 95.266 96.516 97.397 98.018 98.456 98.764 98.982 99.2 </alb>
</length>
<weight>
    <alb>0.061 0.412 1.153 2.244 3.589 5.080 6.624 8.148 9.602 10.953 13.29 15.129 16.526
17.563 18.319 18.866 19.258 19.537 19.736 19.877 20.019</alb>
</weight>

<Mp_mean_max alb="0.15">
    <variable min="0.0" max="0.15" use="false"/>
</Mp_mean_max>

<Mp_mean_exp alb="0.035">
```

-4-

Figure A.4: Example xml parfile (continued)

parfile.xml

```xml
        <variable min="0" max="2.5" use="true"/>
</Mp_mean_exp>


<Ms_mean_max alb="0.01">
        <variable min="0.001" max="1.0" use="false"/>
</Ms_mean_max>


<Ms_mean_slope alb="-0.167">
        <variable min="-0.5" max="0" use="false"/>
</Ms_mean_slope>


<Ms_mean_half alb="34.0">
        <variable min="26" max="100" use="true"/>
</Ms_mean_half>


<M_mean_range alb="0.0">
        <variable min="0" max="1" use="false"/>
</M_mean_range>


<Food_for_age alb="0.9">
        <variable min="0.65" max="1.0" use="false"/>
</Food_for_age>


<forage_ration alb="0.05"/>


<!--========================== -->
<!--FISHERY DATA AND PARAMETERS-->
<!--========================== -->
<file_catch_data>
        <nb_files alb="1"/>
        <alb file1="alb_catch.txt"/>
</file_catch_data>


<file_frq_data>
        <nb_files alb="0"/>
        <alb file1=""/>
</file_frq_data>


<q_sp_fishery>
    <variables use="true"/>
        <L10 alb="3.0e-06">
                <variable min="0.0" max="0.01" use="true" dyn="0.0"/>
        </L10>
</q_sp_fishery>

<!--selectivities as functions of length-->
<!--1 - logistic, 1 parameter: main parameter is used as a slope coefficient-->
<!--2 - sigmoid, 2 parameters: main parameter as slope coefficient + length_threshold-->
<!--3 - asymmetric Gaussian, 3 parameters: length_threshold as optimal value, main as sigma, right_asymptote as the function min on the right -->
<s_sp_fishery>
    <variables use="false"/>
        <L10 alb="0.3">
                <variable min="0.05" max="0.6" use="false"/>
                <function_type value="2"/>
                <length_threshold alb="80.0" use="false"/>
                <right_asymptote alb="0.1" min="0" max="1" use="false"/>
        </L10>
```

-5-

Figure A.5: Example xml parfile (continued)

parfile.xml

```xml
        </L10>
</s_sp_fishery>

<!--================ -->
<!--AGGREGATION ZONES-->
<!--================ -->
<nb_region value="4"/>
<area0 area_id="1" lgmin="130.5" lgmax="165.5" ltmin="-24.0" ltmax="4.0"/>
<area1 area_id="2" lgmin="130.5" lgmax="165.5" ltmin="-54.0" ltmax="-24.0"/>
<area2 area_id="3" lgmin="165.5" lgmax="288.5" ltmin="-24.5" ltmax="4.0"/>
<area3 area_id="4" lgmin="165.5" lgmax="288.5" ltmin="-54.0" ltmax="-24.0"/>

<nb_region_sp_B alb="4"/>
<area_sp_B>
    <alb>1 2 3 4</alb>
</area_sp_B>

<nb_EEZ value="0"/>

</par>
```

Figure A.6: Example xml parfile (continued)

Appendix B

Technical annex

B.1   Makefile

```
HEADERS= \
XMLDocument2.h \
calpop.h \
Map.h \
Matrices.h \
Numfunc.h \
Param.h \
ReadWrite.h \
SaveTimeArea.h \
SeapodymCoupled.h \
SimtunaFunc.h \
StdAfx.h \
mytypes.h \
SeapodymDocConsole.h \
ctrace.h \
Utilities.h \
VarMatrices.h \
VarParamCoupled.h \
VarSimtunaFunc.h

SRCS= \
main.cpp \
XMLDocument2.cpp \
VarParamCoupled.cpp \
seapodym_coupled.cpp \
```

Map.cpp \
Matrices.cpp \
Numfunc.cpp \
Param.cpp \
ReadWrite.cpp \
SaveTimeArea.cpp \
SimtunaFunc.cpp \
VarParamCoupled.cpp \
VarParamCoupled_xinit.cpp \
VarParamCoupled_reset.cpp \
SeapodymCoupled_EditRunCoupled.cpp \
SeapodymCoupled_OnRunCoupled.cpp \
SeapodymCoupled_OnRunFirstStep.cpp \
SeapodymCoupled_OnBuildForage.cpp \
SeapodymCoupled_OnWriteOutput.cpp \
SeapodymDocConsole_UpdateDisplay.cpp \
VarSimtunaFunc_feeding_habitat.cpp \
dv_spawning_habitat.cpp \
dv_juvenile_habitat.cpp \
dv_mortality_sp.cpp \
dv_spawning_habitat.cpp \
dv_caldia.cpp \
dv_tridag_bet.cpp \
dv_calrec_adre.cpp \
dv_survival.cpp \
dv_precalrec.cpp \
dv_calrec_precalrec.cpp \
dv_predicted_catch.cpp \
dv_spawning.cpp \
dv_feeding_habitat.cpp \
dv_total_pop.cpp \
dv_total_mortality_comp.cpp \
dv_starvation_penalty.cpp \
dv_sst_habitat.cpp \
Calpop_caldia.cpp \
Calpop_calrec.cpp \
Calpop_InitCalPop.cpp \
Calpop_precaldia.cpp \
Calpop_precalrec.cpp \
Calpop_tridag.cpp \
Calpop_recompute_coefs.cpp \

like.cpp

SRCPATH=DOM/src:src
INCPATH=-IDOM/src -Isrc
BINPATH=.

OBJPATH=objs
ADMODEL_HOME=/home/isenina/admodel

DEBUG= -g
CFLAGS=-DTRUE=true -DFALSE=false -DVERCONS=1 -D__GNUDOS__ -DOPT_LIB -Wall -Wno-deprecated -I$(ADMODEL_HOME)/include -I/usr/include/libxml2 $(INCPATH)

LFLAGS=-lm -L$(ADMODEL_HOME)/lib -ldf1b2o -ladmod -ladt -lado -ldf1b2o -ladmod -ladt -lado -lstdc++ -lm -lado -ladt -ldl -lxml2
CC=gcc
LL=$(CC)

vpath %.cpp $(SRCPATH)
vpath %.h $(SRCPATH)

OBJECTS=$(SRCS:%.cpp=$(OBJPATH)/%.o)

export OBJECTS
export OBJPATH

export CFLAGS
export LFLAGS
export CC
export LL

all: init $(BINPATH)/seapodym

init:
    @test -d $(OBJPATH) || mkdir -v $(OBJPATH)

test: init $(OBJECTS)
    make -f Makefile.test

docs: $(SRCS) $(HEADERS)
    @doxygen

```
$(BINPATH)/seapodym : $(OBJECTS)
    $(LL) -o$@ $(DEBUG) $^ $(LFLAGS)

$(OBJPATH)/%.o : %.cpp
$(CC) -o$@ $(DEBUG) -c $(CFLAGS) $(filter %.cpp, $^)

$(OBJECTS) : $(HEADERS)

clean:
    @rm -vf $(OBJECTS) $(BINPATH)/seapodym
    @rm -vf $(OBJECTS) $(BINPATH)/seapodym.exe
    @rm -vf $(OBJECTS) $(BINPATH)/TestSeapodymCoupled
    @rm -vrf $(OBJPATH)
    @rm -vrf docs
    @rm -vf gmon.out
```

B.2   Files formats

B.2.1   ASCII fishing data files

The tab delimited text files contaning spatially distributed fishing (Effort and Catch) data have the following structure:

```
Nb of fisheries in the file
Nb of records by fishery  Data Nb by fishery 1  ...  Data Nb by fishery N
f yr mm dd gr lat lon res E C
```

Note that the spatial coordinates (`lon` and `lat`) are given for the center of the region with the resolution given in the `res` column.

B.2.2   Files .bin

These files are produced with the software GMB that redistribute data at the resolution of the selected grid (cf GMB readme file), taking into account the land mask. To get them GMB requires a short configuration file with the paths to original CES files and options for data extraction. The example of such configuration file is given below:

```
6        : number of fisheries files
4        : number of species per file
skj 1    : species and flag to output data for it or not
yft 0    : the second species and flag
bet 0    : ...
alb 0    : the last species and flag
0        : 1 - input files are binary, 0 - ASCII
1        : 0 - output if nonzero catches, 1 - output if nonzero effort, 2 - output :
1        : 1 - output file is binary, 0 - ASCII
path to file1.txt
path to file2.txt
path to file3.txt
path to file4.txt
path to file5.txt
path to file6.txt
```

As the result GMB will create in data directiory six binary files (named as `file*_skj.bin`) with redistributed effort and skipjack catch, if the effort is non null.

Binary format, which is read by Seapodym is similar to ASCII format but containing explicitely the binary addresses to each fisheries time series:
```
First year  First month
Last year   Last month
```

```
Number of species in file
MinLon  MaxLon  MinLat(upper)  MaxLat  : model domain coordinates
MinLon  MaxLon  MinLat(upper)  MaxLat  : fishing area coordinates
Address of first record    Address of last record for time[0]
..
..
Address of first record    Address of last record for time[Tmax]

i  j  effort(i,j)  catch(species[0],i,j)  ...  catch(species[max],i,j)
```

Appendix C

References

Bibliography

[Arditi *et al.*, 2001] Arditi, R., Tyutyunov, Yu., Morgulis, A. Govorukhin, V., Senina, I. 2001. Directed move-ment of predators and the emergence of density-dependence in predator-prey models. *Theoretical Population Biology.* 59(3). 207-221.

[Arditi and Ginzburg, 1989] Arditi, R., Ginzburg, L. R. 1989. Coupling in predator-prey dynamics: ratio-dependence, J. Theor. Biol. 139, 311-326.

[Bard, 1974] Bard, Y. 1974. Nonlinear parameter estimation. Academic Press, New York.

[Barth and Wunsch, 1990] Barth, N., Wunsch, C. 1990. Oceanographic experimant design by simulated annealing. *J. of Phys. Oceanogr..* 20, 1249-1263.

[Berezovskaya *et al.*, 1999] Berezovskaya, F. S., Davydova, N. V., Isaev, A. S., Karev, G. P., Khlebopros, R. G. 1999. Migration waves and the spatial dynamics of phytophagous insects. *Siber. Ecol. J.* 4, 347-357. [In Russian]

[Berezovskaya and Karev, 1999] Berezovskaya, F. S., Karev, G. P. 1999. Bifurcations of running waves in population models with taxis. *Usp. Fiz. Nauk.* 9, 1011-1024. [In Russian]

[Bertignac *et al.*, 1998] Bertignac, M., Lehodey, P. Hampton, J. 1998. A spatial population dynamics simulation model of tropical tunas using a habitat index based on environmental parameters. *Fish. Oceanogr.* 7: 3/4, 326-334.

[Brill, 1994] Brill, R. 1994. A review of temperature and oxygen tolerance studies of tunas pertinent to fisheries oceanography, movement models and stock assessments. *Fisheries Oceanography* 3:3, 204-216.

[Cayre, 1991] Cayre, P. 1991 Behavior of yellowfin tuna (Thunnus albacares) and skipjack tuna (Katsuwonus pelamis) around fish aggregating devices (FADs) in the Comoros Islands as determined by ultrasonic tagging. Aquat Living Resour 4:112

[Chen *et al.*, 1994] Chen, D., Rothstein, L.M., Busalacchi, A.J., 1994. A hybrid vertical mixing scheme and its application to tropical ocean models. Journal of Physical Oceanography 24, 21562179.

[Christian *et al.*, 2002] Christian, J., M. Verschell, R. Murtugudde, A. Busalacchi, McClain, C. 2002. Biogeochemical modeling of the tropical Pacific Ocean I: Seasonal and interannual variability. *Deep Sea Res.* II, 49: 509-543.

[Christian and Murtugudde, 2003] Christian, J., and R. Murtugudde. 2003. Tropical Atlantic variability in a coupled physical-biogeochemical ocean model. Deep Sea Research II, 50(22-26): 2947-2969.

[Czaran, 1998] Czaran, T. 1998. Spatiotemporal Models of Population and Community Dynamics. Chapman and Hall, London.

[Edelstein-Keshet, 1988] Edelstein-Keshet, L. 1988. Mathematical Models in Biology. Mc-GrawHill, New York.

[Faugeras and Maury, 2005] Faugeras, B. Maury, O. 2005. An advection-diffusion-reaction population dynamics model combined with a statistical parameter estimation procedure: application to the Indian skipjack tuna fishery. *Mathematical biosciences and engineering* 2: 4, 1-23.

[Govorukhin *et al.*, 2000] Govorukhin, V. N., Morgulis, A. B., Tyutyunov, Yu. V. 2000. Slow taxis in a predator-prey model, Dokl. Math. 61, 420-422. Translated from Dokl. Akad. Nauk 372, 730-732.

[Griewank and Corliss, 1991] Griewank, A., Corliss, G.F. 1991. Automatic differentiation of algorithms: theory, implementation, and application. SIAM. Philadelphia.

[Grunbaum, 1994] Grunbaum, D. 1994. Translating stochastic density-dependent individual behavior with sensory constraints. Mathematical Models in Biology. McGrawHill, New York.

[Keller and Segel, 1971] Keller, E., Segel, L. A. 1971. Travelling bands of chemotactic bacteria: a theoretical analysis. *J. Theor. Biol.* 30, 235-248.

[Langley *et al.*, 2005] Langley, A., Hampton, J., Ogura, M. 2005. Stock assessment of skipjack tuna in the western and central Pacific Ocean. Western And Central Pacific Fisheries Commission, Scientific Committee,SA WP4). http://wcpfc.org/sc1/pdf/SC1_SA_WP_4.pdf

[Lehodey *et al.*, 1997] Lehodey, P., Bertignac, M., Hampton, J., Lewis, A., Picaut, J. 1997. El Niño Southern Oscillation and tuna in the western Pacific. *Letters to Nature.* 389, 715-718.

[Lehodey, 2001] Lehodey, P. 2001. The pelagic ecosystem of the tropical Pacific Ocean: dynamic spatial modeling and biological consequences of ENSO. *Progress in Oceanography.* 49, 439-468.

[Lehodey *et al.*, 2003] Lehodey, P., Chai, F., Hampton, J. 2003. Modelling climate-related variability of tuna populations from a coupled ocean biogeochemical-populations dynamics model. *Fish. Oceanogr.* 12: 4/5, 483-494.

[Lehodey, 2004a] Lehodey, P. 2004a. A Spatial Ecosystem And Populations Dynamics Model (SEAPODYM) for tuna and associated oceanic top-predator species: Part I Lower and intermediate trophic components. 17th meeting of the Standing Committee on Tuna and Billfish, Majuro, Republic of Marshall Islands, 9-18 Aug. 2004, Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia, Working Paper: ECO-1: 26 pp. http://www.spc.int/OceanFish/Html/SCTB/SCTB17/ECO-1.pdf

[2004b] Lehodey, P. 2004b. A Spatial Ecosystem And Populations Dynamics Model (SEAPODYM) for tuna and associated oceanic top-predator species: Part II Tuna populations and fisheries. 17th meeting of the Standing Committee on Tuna and Billfish, Majuro, Republic of Marshall Islands, 9-18 Aug. 2004, Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia, Working Paper: ECO-2: 36 pp. http://www.spc.int/OceanFish/Html/SCTB/SCTB17/ECO-2.pdf

[Malte *et al.*, 2004] Malte, H., Larsen C., Musyl M.K. and R. Brill. 2004. Differential heating and cooling rates in bigeye tuna (Thunnus obesus): Options for cardiovascular adjustments. Abstract. Comp. Biochem. Physiol. A. 137: S51.

[Matear, 1995] Matear, R. J. 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *Journal of Marine Research.* 53, 571-607.

[McKinley *et al.*, 2006] McKinley, G., et al. 2006. North Pacific carbon cycle response to climate variability on seasonal to decadal timescales. Journal of Geophysical Research, C7, C07S06.

[Morel and Berthon, 1989] Morel, A, J-F Berthon. 1989. Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. *Limnol. Oceanogr.*, Volume 34: 1545-1562.

[Murray, 1989] Murray, J.D. Mathematical Biology. B.: Springer-Verlag, 1989. 767 p.

[Murtugudde *et al.*, 1996] Murtugudde, R., Seager, R., Busalacchi, A.J., 1996. Simulation of the tropical oceans with an ocean GCM coupled to an atmospheric mixed-layer model. Journal of Climate 9, 17951815.

[Nihira, 1996] Nihira, A. 1996. Studies on the behavioral ecology and physiology of migratory fish schools of skipjack tuna (Katsuwonus pelamis) in the oceanic frontal area. *Bull. Tohoku Natl. Fish. Res. Inst.* 58, 137-233.

[Okubo, 1980] Okubo, A. 1980. Diffusion and ecological problems: mathematical models. Springer-Verlag, NY.

[Otter Research Ltd, 1994] Otter Research Ltd. 1994. Autodif: a C++ array extension with automatic differentiation for use in nonlinear modeling and statistics. Otter Research Ltd: Nanaimo, Canada.

[Petrovskii and Li, 2001] Petrovskii, S., Li, B. 2001. Increased coupling between subpopulatons in a spatially structured environment can lead to populational outbreaks. *J. theor. Biol.* 212, 549-562.

[Petrovskii *et al*, 2002] Petrovskii, S., Morozov, A., Venturino, E. 2002. Allee effect makes possible patchy invasion in a predator-prey system. *Ecology Letters* 5 (3), 345-352.

[Robinson and Lermusiaux, 2002] Robinson, A.R., Lermusiaux, P. F. J. 2002. Data assimilation for modeling and predicting coupled physical biological interactions in the sea. From *The Sea*, Volume 12, edited by Allan R. Robinson, James J. McCarthy, and Brian J. Rothschild. John Wiley & Sons, Inc., New York. 475-536.

[Senina *et al.*, 1999] Senina, I., Tyutyunov Yu., Arditi, R. 1999. Extinction risk assessment and optimal harvesting of anchovy and sprat in the Azov Sea. *Journal of Applied Ecology* 36, 297-306.

[Sibert *et al.*, 1999] Sibert, J.R., Hampton, J., Fournier, D.A., Bills, P.J. 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Can. J. Fish. Aquat. Sci.* 56, 925-938.

[Sibert *et al.*, 2006] Sibert, J., Hampton, J., Kleiber, P., Maunder, M. 2006. Biomass, Size, and Trophic Status of Top Predators in the Pacific Ocean. Science 314: 1773-1776.

[Sibert and Hampton, 2003] Sibert, J., Hampton, J. 2003. Mobility of tropical tunas and the implications for fishery management. *Marine Policy* 27: 87-95.

[Skellam, 1951] Skellam, J.G. 1951. Random dispersal in theoretical populations. Biometrika 38, 196-218.

[Turchin, 1998] Turchin, P. 1998. Quantitative analysis of movement. Sinauer, Sunderland, MS.

[Tyutyunov *et al.*, 2004] Tyutyunov, Yu., Senina, I., Arditi, R. 2004. Clustering due to acceleration in the response to population gradient: a simple self-organization model. *The American Naturalist.* 164(6).

[Vallino, 2000] Vallino, J.J. 2000. Improving marine ecisystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research.* 58, 117-164.

[Wang *et al.*, 2005] Wang, X., Christian, J.R., Murtugudde, R., Busalacchi, A.J. 2005. Ecosystem dynamics and export production in the central and eastern equatorial Pacific: A modeling study of impact of ENSO. Geophysical Research Letter, 32, L2608.

[Wang *et al.*, 2006] Wang, X., Christian, J.R., Murtugudde, R., Busalacchi, A.J. 2006. Spatial and temporal variability in new production in the equatorial Pacific during 1980-2003: physical and biogeochemical controls. Deep Sea Research II, 53: 677-697.

[Worley, 1991] Worley, B. 1991. Experience with the forward and reverse mode of GRESS in contaminent transport modeling and other applications. In Griewank, A. and G.F. Corliss. Automatic differentiation of algorithms: theory, practice and application. SIAM, Philadelphia.