Sensitivity analysis and the calibration problem of a biodynamic model for Indian Ocean tuna growth

Nicolas Bousquet[1] & Emmanuel Chassot[2] & Sébastien Da Veiga[3] & Thierry Klein[1] & Bertrand Iooss[1] & Agnès Lagnoux[1]

Institut de Mathématique de Toulouse[1] *& Institut de Recherche pour le Développement*[2] *& SAFRAN Tech*[3]*, France*

The growth of Indian Ocean Yellowfin (*Thunnus albacares*) tunas is characterized by several shapes (Figure 1) which can be explained by the juxtaposition of several *environmental forcing* parameters $X \in \Omega \subset \mathbb{R}^d$ (for instance the water temperature or the food density) and *metabolic* (intrinsic) parameters $\theta \in \Theta \subset \mathbb{R}^q$. Several regime shifts can appear, that are connected to specific development stages: typically, larvae grow fast to juveniles, then some bioenergetic losses may occur, indicating spawning or senescence. Nonetheless, older and healthy fish are experimented predators, which may be traduced by a sensible increase of growth acceleration towards an asymptotic limit, in favorable conditions. The knowledge of the values range of the most influential parameters driving each kind of shape and the age-length key also produced can play an important role in the determination of the size structure of fishing gears, in a perspective of elaborating sustainable exploitation patterns.

Calibrating and classifying the range of values for (X, θ), in function of shapes, can be conducted using a biodynamical, functional computer model $\{L(t), W(t)\}_{t=0,\ldots,T} = g_\theta(X)$ based on the Dynamic Energy Budget (DEB) theory [1]. It simulates simultaneously the curves of fork length $L(t)$ and the weight $W(t)$ indexed by age $0 \leq t \leq T$, and several sources of information. The latter are described as pointwise noisy observations $\mathbf{D}^*$ of fork lengths and weights arising from laboratory experiments, commercial catches and capture-recapture campaigns (Figure 2). More formally, the aims are:

1. to start from two prior distributions $f_X(x)$ and $\pi(\theta)$, typically based on previous works on close species, and to conduct a first sensitivity analysis (SA) to highlight the most influent inputs and decrease the dimension ($d+q \geq 20$); classical SA tools (differential SA, multivariate Sobol' indices [2], etc., see [6] for a review) and others (e.g., elasticity indices) are used to do so, since no dependence between the inputs is assumed a priori;

2. to update $f_X(x)$ and $\pi(\theta)$ conditionally to the likelihood $\ell(\mathbf{D}^*|X, \theta)$ by computing the posterior distribution with density

$$h(x, \theta|\mathbf{D}^*) \quad \propto \quad \ell(\mathbf{D}^*|x, \theta) f(x) \pi(\theta); \tag{1}$$

3. to conduct a more detailed sensitivity study, taking into account the classification of shapes arising from the knowledge of (1) and the correlation between the inputs. More elaborated SA approach are required, as in [3].

The Bayesian computation of the posterior faces difficulties linked to the nature of observations: capture-recapture data have correlated noises, while one of the dataset $\mathbf{D_3} \subset \mathbf{D}$ provides a huge number of correlated fork lengths and weights, without age index. Therefore the full likelihood of observations is not tractable.

Hence, in a first step, “likelihood-free” methods as Approximate Bayesian Computation (ABC, [4]) appear to be useful to update the original prior $f_X(x)\pi(\theta)$ in a more informational prior $\tilde{h}_\epsilon(x, \theta) \propto \mathcal{M}\epsilon\{\mathbf{D^*}_3, \mathbf{D}_3(x, \theta)\} f_X(x)\pi(\theta)$, where $\mathcal{M}\{.,.\}$ is a set of similarity measures between true ($\mathbf{D^*}_3$) and simulated ($\mathbf{D}_3(x, \theta)$) observations. It is indexed by a parameter vector ϵ determining a limit value for the similarity. This ABC step is followed by a kernel reconstruction of the new prior $\tilde{h}_\epsilon(x, \theta)$ based on copulas (R-Vines) [5]. This formal prior distribution, indeed, allows to

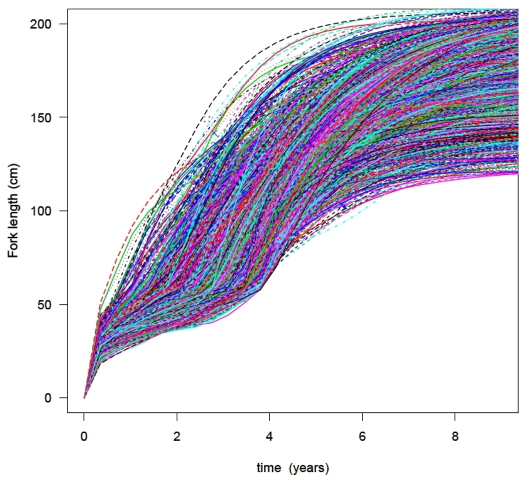

Figure 1: Typical shapes of Indian Ocean Yellowfin growth (fork length versus age).

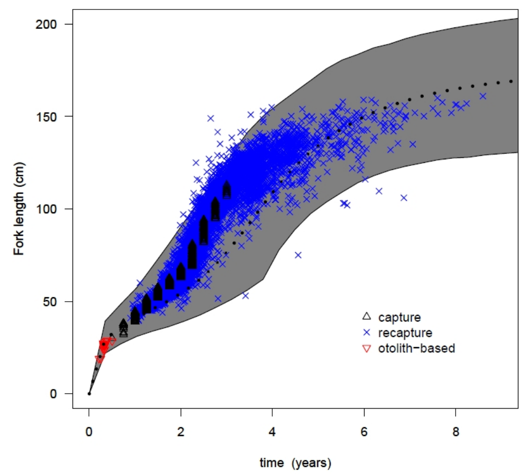

Figure 2: Posterior credibility (90%) area for growth curves arising from the calibrated model (fork length versus age).

develop an adaptive algorithm of posterior calibration based on classic Markov Chain Monte Carlo (MCMC) methods.

The results highlight the strong mixed several influence of environmental and intrinsic parameters, and the calibration methodology developed in this work allows to discriminate the influential factors that explain the variety of shapes. This can apply to a large number of studies involving noisy ground experiments and difficulties to connect correlated observations to functional computer models through statistical likelihoods.

References:

[1] E. Dortel, L. Pecquerie, E. Chassot. "A Dynamic Energy Budget modeling approach to investigate the eco-physiological factors behind the two-stanza growth of yellowfin tuna". *Submitted*, 2016.

[2] F. Gamboa, A. Janon, Th. Klein, A. Lagnoux. "Sensitivity analysis for multidimensional and functional outputs". Electronic Journal of Statistics, 8, 575-603, 2014.

[3] S. Da Veiga. "Global sensitivity analysis with dependence measures". Journal of Statistical and Computational Simulation, 85, 1283-1305, 2015.

[4] J.-M. Marin, P. Pudlo, C.P. Robert, R.J. Ryder. "Approximate Bayesian computational methods". Statistics and Computing, 22, 1167-1180, 2012.

[5] T. Bedford, R. M. Cooke. (2001). "Probability density decomposition for conditionally dependent random variables modeled by vines". Annals of Mathematics and Artificial Intelligence 32, 245-268, 2001.

[6] B. Iooss and P. Lemaître. "A review on global sensitivity analysis methods". In: *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, C. Meloni and G. Dellino (eds.), Springer, 2015