**ICCAT GBYP CORE MODELLING MSE GROUP**

**ICCAT Secretariat, Madrid, Spain**

**19 and 23 July 2017**

**Draft report**

The meeting started with some delay, due to the absence of some members.

At 10:30 it was decided to formally open the meeting, adopting the draft agenda.

The Group decided to nominate Doug Butterworth as Chair and Tom Carruthers as co-Chair.

Antonio Di Natale was identified as rapporteur for the general matters, while Tom Carruthers was identified as the rapporteur for the modelling matters.

The Group welcomed Shuya Nakatsuka as observer/expert for the meeting.

As concerns point 4 of the agenda (Activities to be done in agreement with the proposals set by the CMMG in March 2017 and shared with the ICCAT SCRS Bluefin Tuna Species Group), Butterworth introduced the discussion, with a specific reference to annex 1 of the draft agenda. Several points have been raised about the possible enlargement of the Group with other experts, the opportunity to involve managers in the process, the experiences done so far by other TRFMOs and other Commissions (i.e. IWC), the schedule for delivering the MSE and the OM, etc.

The Group raised the problem of scheduling the second part of the meeting during the BFT Assessment Meeting. According to what was discussed in March, the second part of the CMMG was scheduled for Sunday, but the Group asked Gordoa to agree with the Chair and the WBFT Rapporteur the best opportunity for this second part of the meeting, taking into account the needs of both meetings.

As concerns point 5, Carruthers presented the progress since the last meeting in March.

It was pointed out the importance to use the various sets of GBYP data in the OM and how these data are used.

The Group, revising what it was discussed at the Monterey meeting, decided to make publically accessible the software developed by the MSE expert, using the ICCAT web site. The public can now access the software on [ht tps://github.com/ICCAT/abft-mse/wiki](https://github.com/ICCAT/abft-mse/wiki)

The Group noticed inconsistencies in some specific data series that have been used (area, quarter), pointing out the problems that shall be fixed.

The various scenarios have been discussed and some aspect of the recruitment have been pointed out as problematic, possibly because of the lack of size frequency data for juvenile classes in the recent years.

Several discussions concerned the outlook of the data concerning the two stocks, compared with MSY and the historical level of unfished stock (both Base M and 2/3 Base M). Several doubts have been raised about the representation of the historical catches prior to about 1975 and even more on data prior to 1950. It was pointed out that the representation is more reliable in the last part of the plot, when more robust frequency and fishery data were available. The group decided to come back to the plots when the new assessment will be available, for better checking the assumption.

Looking at the data series that were used for the OM, there were discussions about the data from the WATL PS, particularly for the data coming from the ‘80s, and much more discussions about the EATL BB data, which currently include bot the BB fishery in the Bay of Biscay and the BB fishery from the Canary Islands. Another discussion was concerning the WATL RR data and even there the cleaned catch-at-size data are possibly the issue. The PS series used for the MED were also discussed and it was suggested to possibly split the Adriatic PS from the others, due to different target in size. It was raised the problem that the standardised Sardinian trap data are currently not used in the OM even if they are the only fishery index for the Mediterranean.

The discussion was also addressed to the reference OM factors and levels and particularly in the spawning fraction as it was considered. Another discussion concerned the choice of M, which will be further discussed by the Assessment meeting. It was raised the problem, already faced during the previous meeting, about the different shape in M and it said that it would be quite useful to use the same M used in the assessment, at least for consistency.

The Group checked the reference of the various data sets that are used for conditioning the OM, one by one, checking the correct attribution of quarter and area. Inconsistencies were noted for the SA\_MOR\_TRAPS and for the MOR\_POR\_TRAPS that should be further checked in terms of quarter attribution. Further checks are also needed for the JP\_LL\_EMED, in terms of the area. Further checks will be necessary also for the series labelled as GSL, because of the limited extension of the area.

Item 6. OM conditioning

When conducting Management Strategy Evaluation an important step is the development of the Operating Model (OM), a mathematical–statistical model that is used to describe the resource dynamics in simulation trials and to generate resource monitoring data when projecting forward. Since once the OM and management objectives have been agreed the “best” MP is an emergent property.

The OM has to be conditionied on data, knowledge and hypotheses that reflect uncertainty about the dynamics. This requires developing scenarios for alternative model structures and datasets, validation of the model and then rejection and weighting of the scenarios. Validation of models is normally done by comparing estimates (i.e. predictions) to observations. In fish stock assessment this is normally based on a naive adaptation of Pearson residuals, i.e. the difference between observations and posterior means, even if this approach is flawed.

A reason for this is because statistics based on residuals from model fits are not necessarily, a reliable guide to how well that model will predict, since a high R2 or low root mean square error (RMSE) can be obtained by over fitting. For example, in a simple polynomial regression better fits to the data can be obtained by adding higher order terms but the predictions from the model on new data will usually get worse as higher order terms are added. While if scenarios comprise alternative model structures and datasets then AIC cannot be used to compare them. There are also a number of potential problems to identify when examining residuals, e.g. bias, drift, skewness, missing variables, and heteroscedasticity. When inspecting residual patterns, however, there is a danger of hypothesis fishing and so it is good practice to reserve part of the data for validation. This ensures that the significance of a pattern in the data is not tested on the same data set which suggested the pattern.

Cross validation is a technique for evaluating the predictive error of a model by testing it on a set of data not used in fitting. There is often insufficient data, however, in stock assessment datasets to allow some of it to be kept back for testing. A more sophisticated way to create test datasets is, like the jackknife, to leave out one (or more) observation at a time. Cross validation then allows prediction residuals to be calculated, i.e. the difference between an observation and its out-of-sample predicted value. Model validation examines if the model family should be modified or extended, and is complementary to model selection and hypothesis testing. In contrast model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced.

The OM is used to model dynamic processes, e.g. to account for process error, therefore model residuals correspond to measurement error and process error not accounted for in model conditioning. There is also a danger of over fitting when conditioning the OM on historical data, which will result in failure to project forward reliably.

If the data are regarded as being representative of the dynamics of the resource then they can be used as a model-free validation measure. This allows examination of prediction errors to be used for validation, rejection of models and indicating useful directions in which the model could be improved. Cross-validation can also help to identify which data series are appropriate for use in the MP and for specifying measurement error for generating resource monitoring data by the Observation Error Model (OEM) when projecting forward.

***Proposal***

Run a model free cross-validation and estimate prediction errors for the OM scenarios in order to

* Validate the Oms, i.e. do some scenarios better fit some of the data series than others.
* Evaluate how to extend model structure, in order to better account for process error
* Propose how to develop scenarios based on alternative data weighting schemes and model structures
* Condition the OEM, i.e. by identifying the error structure of CPUE series

After several further discussions, the meeting was adjourned to Friday at 6 pm, for a quick check of the advancements. At the same time, during the following days, a validation of the assumptions should be necessary.

In the last part of the first day of meeting, the Group reviewed the MSE work done by the IWC, and particularly the structure of multi-meeting used for developing the MSE and the OM.