Inter-observer reliability of fish vitality assessments: does the use of categorical or continuous scoring scales affect the Reflex Action Mortality Predictor (RAMP)?

Meeremans Pieter[1,2], Marc Kochzius[2], Noelle Yochum[3] and Sebastian Uhlmann[1]

[1]   Institute for Agricultural and Fisheries Research (ILVO), Ankerstraat 1, 8400 Oostende, Belgium
      E-mail: Pieter.Meeremans@vub.ac.be

[2]   Marine Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

[3]   Oregon State University, 104 Nash Hall, OR 97331 Corvallis, United States of America

To incentivise fishers to fish more selectively, the revised Common Fisheries Policy (CFP) includes an obligation to land all catches of quota-regulated species. From January 1st, 2016, it applies to some demersal fisheries and their target species and by 2019 to catches of all quota-regulated species. To stay viable, those fisheries that struggle to improve selectivity, may seek exemptions from this obligation. One exemption requires scientific evidence to demonstrate "high survival" of discarded fish. Based on international guidelines, methods have been developed in Belgium (Uhlmann et al., in press) to investigate the survival probability of plaice (*Pleuronectes platessa*). This was done using the reflex action mortality predictor (RAMP) approach, which relates status indicators such as vitality and/or injury to mortality probabilities. In other countries, this and similar techniques have been used. However, when several scientific observers with different levels of experience and training become involved in collecting vitality and injury data, there is potential for bias from subjectivity in measurement. To improve credibility of results and maximize accuracy, it is important to assess inter-observer reliability and evaluate whether any variability affects significance (if any) of a vitality-mortality relationship. For the RAMP approach, vitality and injury assessments are completed by scoring impairment in pre-established reflex action behaviour and injury categories. There are multiple ways to calcuate these 'scores', including categorically (i.e., injury/reflex action absent, weak, or present) or continuously (i.e., scale from 0-10). Scoring based on these two metrics may result in different relationships with mortality.

So far, 15 laboratory-held individuals of each plaice and common sole (*Solea solea*) were scored, using both the categorical and continuous metrics, by four different observers. Future trials will involve four days on-board the R/V *Simon Stevin* in February/March 2016, when at least three observers will simultaneously score each of ~400 beam-trawled plaice by using both categorical and continuous scoring scales. All scored plaice will be returned to laboratory holding facilities in Ostend and monitored for any mortality for 14 days. Based on these trials, we will create RAMPs using both categorical and continuous scoring data, and we will evaluate inter-observer variation. Through this analysis, we will be able to determine which RAMP scoring metric is less influenced by observer bias, and whether scores from multiple observers may be used as a proxy for survival (i.e., RAMP) to facilitate implementation of the landing obligation.

References

Uhlmann, S. S., Theunynck, R., Ampe, B., Desender, M., Soetaert, M., and Depestele, J. Injury, reflex impairment, and survival of beam-trawled flatfish. - ICES Journal of Marine Science, doi: 10.1093/icesjms/fsv252.