CTRM Program Redevelopment: Interim Summary Report

Liam Furman and Susan Johnston

Marine Resource Assessment and Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch 7701

February 2010

Introduction:

Fishing trawlers record their catches in two ways, after each individual trawl (called the drag record) and also the totals when discharging at shore (called the landing record). Unfortunately, drag records, which are used by GLMs (generalised linear models), are often unreliable. The essence of the CTRM (convert to real mass) program is to use hake landing records to build more reliable and consistent drag records.

The general process can be outlined as follows. Suppose that the total catch of a given product is reliably reported in the landing record. We look up that product in the drag record to find the proportion of the product caught in each individual trawl. Using this information, the reported landed amount is divided among individual drags, effectively making a new drag record. Unfortunately, there are complications. Different codes are often used for the same products or groups of products at the landing and drag levels. This means that there is no simple one-to-one correspondence between products in both records; however, by inspecting them and applying decision rules, we can make educated guesses as to how best to map between records.

A further complication is that the records often contain errors, such as invalid product codes or impossible discrepancies between the landing and drag records. Therefore, the program also detects errors and reports them so that they can be investigated by the user. This can help to highlight previously unrecognised situations, which require further decision rules to process suitably.

Progress:

The program been coded in Python, an easy to use language, to interact with an MS Access database using SQL. It complies with the latest revision of Rob Leslie's *CTRM Program* document, which deals only with Hake data. The two distinct phases of the program covered by the document have been fully coded, expect for a few points on which clarification is awaited.

Phase one includes pre-processing and error detection. First, the program imports the required records and reference tables from spreadsheet format into a database for ease of use, although it can be directly connected to an existing database if preferred. The program then processes the records to generate landing and drag summary lists. These lists contain most of the useful information, which reduces the amount of SQL and provides programming shortcuts. Finally, they are used to check for errors in the records. The errors include invalid product codes, inconsistencies with regard to vessel type (freezer or wetfish) and invalid raising factors.

Phase two apportions landed catches among drags according to proportions from the drag summary list. This creates a new, more reliable drag record which uses fewer category codes by grouping similar products together. The processing is broken down according to vessel type and then by

product category (e.g. H&G Wetfish). Each section uses decision rules to decide how the mapping should proceed. This often amounts to checking for the presence or lack of certain category codes in the drag and landing records to suggest how the codes were being used. Errors can be found at this stage by picking up on discrepancies between records.

All errors are written to an error log with helpful messages, in the expectation that commonly arising problems can be addressed by further decision rules so that more data can be included in the files developed for eventual analysis.

Findings:

So far, by running the program with only a limited amount of test data, two new errors have been discovered. First, sometimes products are reported on the landing record before those products have valid raising factors. Second, in the records some vessels are marked as being both wetfish vessels and freezer vessels on the same trip.

Remaining work:

The few points from the *CTRM Program* document that are awaiting clarification need to be discussed with Rob Leslie and then incorporated into the program.

Although most of the logic has been coded, the program still needs to be optimised. There is a large amount of data that needs to be processed and any small improvement could substantially decrease the running time of the program.

The program needs to be robustly tested to minimise bugs. Owing to the large amount and nature of the data it is unlikely that bugs will be picked up easily. It is probably necessary to design a test module if the program is to be used with confidence.

At this time the program has only been run with a small amount of ‘play’ data. A full decade or so of data must be processed to check for bugs, but also to find unexpected mapping situations and errors in the records. The solutions, probably in the form of further decision rules developed in consultation with MCM staff, will then need to be incorporated into the program.

A full technical specification must be written describing the program’s usage and its algorithms.

Acknowledgments

Rob Leslie and Tracey Fairweather of MCM are thanked for assistance with queries.