The three Vs of tag recovery data processing: validation, validation, validation

A just-tagged yellowfin tuna is measured before going back to the ocean (image: Bruno Leroy, SPC).

The Pacific Community (SPC) has tagged and released tunas in the western and central Pacific Ocean since 1977. These tagging efforts have been carried out through three tagging programmes: the Skipjack Survey and Assessment Programme from 1977 to 1981; the Regional Tuna Tagging Programme, from 1989 to 1992; and, the current Pacific Tuna Tagging Programme (PTTP), since 2006. In total, more than 700,000 tuna have been tagged and released, of which 100,000 have been recovered and reported to SPC. Tagging experiments provide important inputs to analyses that support management of tuna resources in the region, including stock assessment models.

Tagged fish can be detected months after recapture and at a range of stages in the tuna supply chain, for example: on fishing vessels during well transfers or sorting; during the transfer of tuna from fishing to transhipment vessels, and unloading from transhipment vessels; in cold-storage facilities; and in canneries. Previous studies have demonstrated relationships between the reliability of tag recovery information and other variables, such as the time elapsed between tag recovery and tag detection. As such, a critically important component of the tagging programmes is rigorous data quality control by SPC staff, including cross validation of tag recovery data with available information from other datasets. After cross-validation, each recovery is assigned a best estimate of recovery date and position, and reliability indices for the perceived uncertainty in these key data fields (e.g. a recapture date best estimate of 12 March 2015, ± one week). Cross-validation compares tag recovery information from external datasets such as vessel tracks from vessel monitoring system (VMS) data and vessel logbook data, in order to estimate the reliability of information provided by tag finders. For example, VMS data could be used to confirm that the fishing vessel reported to have recovered a tag was likely fishing in the vicinity of the reported recovery position on the reported tag recovery date, based on the location and speed and azimuth profile of the vessel. Additional controls based on the growth rate of the tuna during its time at liberty, the maximum distance travelled or the bathymetry are also used to ensure the quality of the data. The cross-validation process represents a considerable investment of time and resources, particularly when high numbers of tag recoveries are reported shortly after tagging cruises, with up to four staff contributing on a part-time basis.

However, this poses an important question: How accurate are the perceived reliability indices for recovery date and position? And how can we determine this, if we do not know exactly when and where tags are recaptured by fishing vessels? Tag seeding experiments provide the means to investigate these questions. In tag seeding experiments, observers on purse seiners secretly insert conventional tags in tuna. Recovery data from tag seeding experiments receives the same data quality controls as PTTP recoveries. The date and location of seeding can then be compared directly with the best estimates of recovery date and position resulting from the cross-validation process, and the errors in the best estimates compared with the perceived reliability assigned during cross-validation.

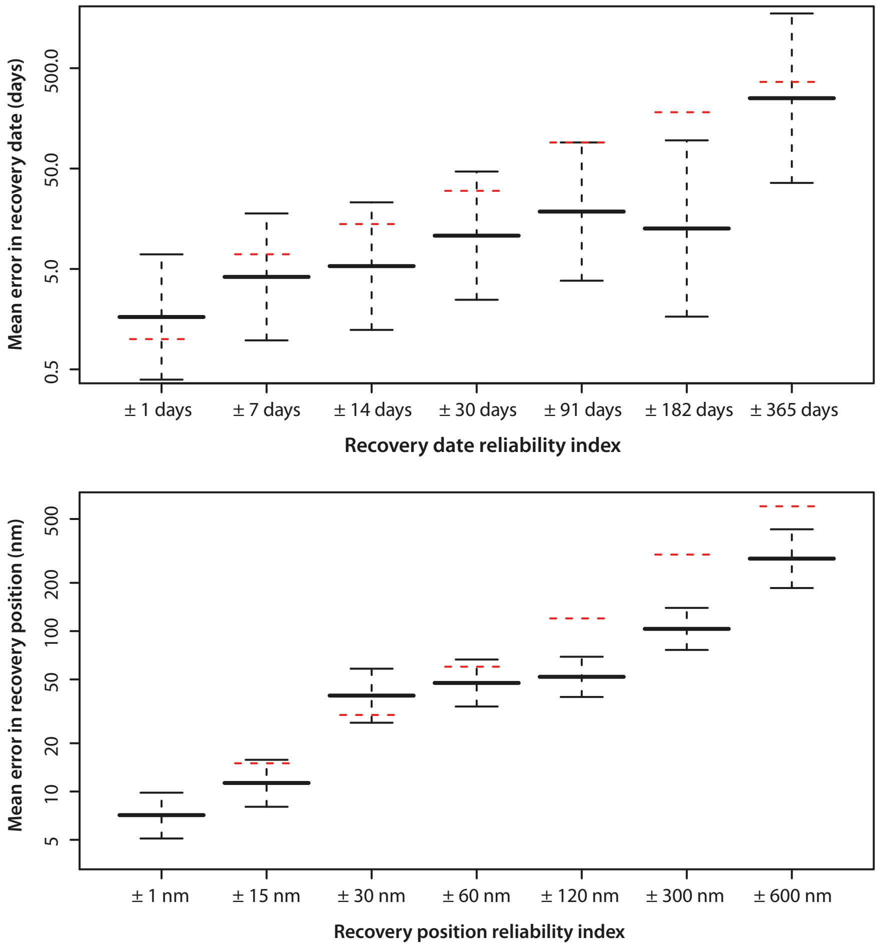

Figure 1. Estimated mean errors in recovery date (top) and position (bottom) against assigned reliability indices. Thick black lines give the best estimate, and thin black lines display uncertainty in the best estimate. Dotted red lines display the perceived accuracy from the cross-validation process.

The errors in recapture date and position for seeded tags increase in proportion to their perceived uncertainty (Fig. 1). This demonstrates that the cross-validation process is capable of accurately determining the relative reliability of tag recovery information. For tag recoveries with high perceived uncertainty in recovery information, the best estimates of recovery date and position are often more reliable than suggested from the cross-validation process. In general, however, the observed errors in recovery position and date are consistent with the perceived uncertainty from the cross-validation process. This suggests that the cross-validation process is also capable of estimating the magnitude of uncertainty in recapture date and position.

Tag seeding information also allows the exploration of factors that influence the accuracy of estimated uncertainty in recovery date and position. For example, the majority of checks and comparisons undertaken during cross-validation rely on accurate identification of the tag recovery fishing vessel. Unsurprisingly, there is clear evidence that errors in recovery date and position are higher when recovery fishing vessels were not correctly identified by tag finders.

The analysis summarised here provides encouraging evidence that the cross-validation process provides accurate estimates of uncertainty in tag recovery information. This allows uncertainty in tag recovery information to be accounted for in analyses of tagging data. Furthermore, tag seeding data provides an opportunity to explore the factors that have the most impact on uncertainty in recovery information. This can inform changes to the tagging programme that can improve confidence in the resulting tagging dataset.

For more information:

Thomas Peatman
Senior Fisheries Scientist (Statistical Modelling), SPC
thomasp@spc.int