Match My Whale: A crowdsourcing platform for cross-identification of Australian humpback whales (*Megaptera novaeangliae*)

Stack, S.H., Swabb, M.H., Currie, J.J., Kaufman, G.D., Martinez, E.

Pacific Whale Foundation, 300 Ma'alaea Road, Suite 211, Wailuku, Maui, HI 96793, USA

Abstract

The large size of Australian humpback whale (*Megaptera novaeangliae*) photo-identification (ID) catalogs is both a strength and a challenge. Every year, researchers add hundreds of new flukes photographs to their catalogs, making the task of pairwise checking all individuals exponentially more difficult. Large catalogs mean that cross-checking is prohibitively time consuming, and, furthermore, within a catalog more and more errors can accumulate during the matching process (specifically, "false negatives" from missed matches) which reduces the confidence in the data and statistical models (Stevick *et al.*, 2001). These missed matches can result in overestimates in abundance and growth because previously sighted individuals are miss-categorized as new individuals (Hammond *et al.*, 1990).

A new website "Match My Whale" has been developed by Pacific Whale Foundation (PWF) to harness the power of crowd-sourcing and test the theory that this online citizen scientist fluke matching platform will be more effective than the current method(s) of manually searching for a match, or relying on complicated computer software. PWF considers this to be a breakthrough method to manage thousands of individual photos and get the most accurate data for sight-resight models and provide insight into humpback whale population abundance and distribution. Crowdsourcing can offer the effort and redundancy needed for scientists to manage their current catalogs, and facilitate the integration of multiple catalogs from different organizations.

Keywords: Crowdsourcing, Citizen Science, Humpback Whale, Photo-ID, *Megaptera novaeangliae*, MyWhale.org, Match My Whale.

Introduction

In 2013, a Drupal 7 CMS-based development website was created. The appropriate domains were obtained and in February 2014 a "coming soon" page was launched. PWF also obtained the relevant social media counterparts for Match My Whale (eg. Facebook, Twitter, Instagram) and in May 2014 will use these social media sites, as well as PWF's own social media sites and email lists, to promote the "coming soon" site.

The newly created web application has appeal messages and content about the project, as well as humpback whale fluke-ID activities. A complete data model for the application has been created, which is the structure of the data relationships and a representation of the data being collecting. Two photo catalogs were combined for the initial launch: one from PWF in East Australia and one from the Centre for Whale Research (CWR) in West Australia. The large initial data set,

consisting of 8,047 photographs, of which 80% came from PWF and 20% from CWR, was cropped and re-sized to the desired specifications, and uploaded to an online server, which will host the web application. The user interface, web pages, features, and algorithms have been designed. Tutorials were created so that citizen scientists are properly trained in the three "research activities" of SPLASH scoring (Calambokidis *et al.*, 2001), fluke classification, and fluke matching. Users must pass a test in each of these three categories before they can participate in the actual activity. Citizen scientists may choose to participate in any one, two, or all three of these activities, but they must successfully pass the test before being allowed access to the cataloged images. PWF is currently alpha-testing the website internally and will start beta-testing with members of the public over the summer 2014.

The final feature of the website to be added is the ability for the public to upload their own humpback whale fluke photographs. This feature should be completed in June 2014. Only registered users will be able to upload photos to the site. Some photography tips and recommendations on the upload page will also be provided in hopes of receiving only high quality photographs. In addition, users will be recommend to take the SPLASH scoring test to understand the photo quality required in a photo-ID photograph, but this is not a requirement.

The Match My Whale website provides:

- A user friendly interface to help guide the citizen scientist through the steps of SPLASH scoring, classifying the fluke based on pigmentation, and searching for a match.

-A side-by-side display that allows the user to compare photo-IDs when looking for fluke matches.

-The ability for users to upload their own humpback fluke photographs, and match them once these have been SPLASH-scored by a third party.

Methods and Objectives

This project aims to substantially improve the manual matching methods currently in use by researchers by harnessing the power of citizen scientists. PWF feels this project is incredibly appropriate, as many research institutions worldwide have humpback whale catalogs but have limited resources and man-power to keep those up-to-date and/or integrate multiple catalogs. The use of different software available, such as Fluke Matcher (Kniest *et al.*, 2010) and Discovery (Gailey and Karczmarski, 2011), can also make the cross-matching process between catalogs quite challenging among researchers. Bringing these photographs to an online platform solves the availability and consistency in the database issues, while having citizen scientists donate their time to score photo-IDs and/or search for matches solves the problem of scientists not having enough time or personnel to do the matching themselves. Another advantage of this project is the ability of users to upload their own photos, which not only increases the amount of data available to researchers but can also provide further information in remote and/or less

studied areas. Finally, this project has an important social aspect as it also involves the public in humpback whale research, instilling a sense of ownership and protectiveness of the whales, which is reflected in the domain name "my whale".

Upon proving the efficacy of Match My Whale, PWF will seek to include other major Australia data sets, collectively representing an additional > 10,000 individual ID photos. PWF is also seeking collaboration with other institutions who may be interested in donating their humpback whale fluke catalogs to Match My Whale. PWF welcomes all who wish to partner with us by sharing their catalogs and encourage them to contact us at catalog@mywhale.org. At this time, PWF is only accepting fluke images taken in Australian waters, but PWF's vision is to expand this project to accepting fluke photos of the North Pacific humpback whales taken in Hawaiian waters, and then worldwide.

Once launched, PWF will be measuring the effectiveness of crowdsourcing as an efficient cross-catalog matching mechanism. If proven effective, the integration of other catalogs as well as photo-IDs collected by the public will help improve our knowledge on the life histories of the Australian humpback whale populations. Match My Whale has the potential to be applied to other humpback whale populations around the world, as well as other cetacean species.

Timeline of Project Outputs

Milestone Description	Anticipated Date(s)
Feature enhancement and alpha testing phase	May-June 2014
Beta public launch (limited number of users)	July-2014
Ability for public to upload their own image (East and West Australia only)	July-2014
Full public launch	November-2014
Ability for public to upload their own image (expansion to include North Pacific humpback population)	December-2014
Utilization by IWC Scientific Committee	May-2015
Ability for public to upload images worldwide	December-2015

Management Outcomes

Having high quality photo-identification data and cross-checking of ID catalogs can benefit ongoing comprehensive stock assessment studies for the Southern Hemisphere humpback whales. This project will indirectly benefit the models to estimate survival probability, re-sight probability, and population rate of increase (Pradel, 1996; Petit and Valiere, 2006). Should crowdsourcing prove to be an efficient way of matching individuals across catalogs, it may also provide empirical estimates of emigration and transience, rather than estimates inferred from sight-resight models (Pradel *et al.*, 1997; Fujiwara and Caswell, 2002).

Upon the proven success of this website, PWF will be soliciting the IWC to act as a curator for a potentially very large fluke image repository. The IWC has therefore the potential to host a worldwide catalog of humpback whale photo-IDs, accessible to research groups.

Acknowledgements

References

Calambokidis, J., Steiger, G.H., Straley, J.M., Herman, L.M., Cerchio, S., Salden, D.R., Urbán R., J., Jacobson, J.K., von Ziegesar, O., Balcomb, K.C., Gabrielle, C.M., Dahlheim, M.E., Uchida, S., Ellis, G., Miyamura, Y., Ladrón de Guevara P., P., Yamaguchi, M., Sato, F., Mizroch, S.A., Schlender, L., Rasmussen, K., Barlow, J. and Quinn II, T.J. 2001. Movements and population structure of humpback whales in the North Pacific. *Marine Mammal Science* 17 (4):769-794.

Fujiwara, M. and Caswell, H. 2002. A general approach to temporary emigration in mark-recapture analysis. *Ecology* 83(12): 3266-3275.

Gailey, G. and Karczmarski, L. 2011. Discovery: Photo-identification data-management system for individually recognizable animals. Available online: http://www.biosch.hku.hk/ecology/staffhp/lk/Discovery/

Hammond, H.S., Mizroch, S.A., and Donovan, G.P. 1990. Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters. *Report of the International Whaling Commission Special Issue,* 12. Cambridge. 440 pp.

Kniest, E., Burns, D. and Harrison, P. 2010. Fluke Matcher: a computer-aided matching system for humpback whale (*Megaptera novaeangliae*) flukes. *Marine Mammal Science* 26(3): 744-756.

Petit, E. and Valiere, N. 2006. Estimating Population Size with Noninvasive Capture-Mark-Recapture Data. *Conservation Biology* 20: 1062-1073.

Pradel, R. 1996. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics* 52(2): 703-709.

Pradel, R., Hines, J.E., Lebreton, J.D. and Nichols, J.D. 1997. Capture-recapture survival models taking account of transients. *Biometrics* 53: 60-72.

Stevick, P.T., Palsboll, P.J., Smith, T.D., Bravington, M.V., and Hammond, P.S. 2001. Errors in identification using natural markings: rates, sources, and effects on capture–recapture estimates of abundance. *Canadian Journal of Fisheries Aquatic Science* 58: 1861–1870.