Epigenetic Age Estimation for Hawaiian False Killer Whales (*Pseudorca crassidens*) in the Absence of 'Known Age' Individuals

Karen Martien[1], Robin Baird[2], Kelly Robertson[3], Michaela Kratofil[4], Sabre Mahaffy[5], Kristi West[6], Susan Chivers[7], Frederick (Eric) Archer[8]

[1]NOAA Southwest Fisheries Science Center, La Jolla, CA, [2]Cascadia Research Collective, Olympia, WA, [3]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, La Jolla, [4]Cascadia Research Collective, Oregon State University, [5]Cascadia Research Collective, Olympia, Washington, [6]Kailua, HI, [7]Southwest Fisheries Science Center, La Jolla, CA, [8]Southwest Fisheries Science Center, La Jolla, California

Epigenetic aging models hold great promise for enhancing many aspects of wildlife research and management. However, their utility is limited by the need to train models using known-aged animals, which are rare among cetaceans. We developed and evaluated methods to train an epigenetic age model for Hawaiian false killer whales (*Pseudorca crassidens*) using 94 samples from 78 individuals with ages estimated from sighting histories, morphology, and associated data. Each age estimate was assigned a confidence rating from 1-5 (low-high). We evaluated three different analytical frameworks: elastic net regression, Random Forest (RF) regression, and Bayesian modeling. For the regression methods, we used point estimates of age weighted by their confidence rating. For the Bayesian model, we used age probability distributions for each sample. The median age error (MAE) for samples with high confidence ratings was lower for RF regression (2.57 years) than for elastic net regression (3.63 years). The RF regression MAE is comparable to the accuracy of published cetacean epigenetic clock models trained using individuals with high precision known ages. RF regression was also more consistent across replicate runs than elastic net regression in terms of MAE and the number of predictors included in the final model. The modes of the age estimates from the Bayesian model are less accurate than for RF regression. However, the Bayesian model has the advantage of producing full posterior age probability distributions for each sample, rather than a single point estimate of predicted age. The Bayesian model also accounts for uncertainty in estimates of methylation that result from low sequencing coverage. We caution against relying on MAE as the primary metric of model accuracy because age errors for individual samples can be considerably larger than the median, and recommend further development of the Bayesian model to facilitate incorporating epigenetic estimates into population models.