A right whale pootree: classification trees of faecal hormones identify reproductive states in North Atlantic right whales (*Eubalaena glacialis*)

Peter Corkeron[1,*], Rosalind M. Rolland[2], Kathleen E. Hunt[2] and Scott D. Kraus[2]

[1]National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA
[2]Anderson Cabot Center for Ocean Life, John H. Prescott Marine Laboratory, New England Aquarium, Boston, MA 02110, USA

*Corresponding author: National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA. Tel: +1 508 495 2191. Email: peter.corkeron@noaa.gov

Immunoassay of hormone metabolites extracted from faecal samples of free-ranging large whales can provide biologically relevant information on reproductive state and stress responses. North Atlantic right whales (*Eubalaena glacialis* Müller 1776) are an ideal model for testing the conservation value of faecal metabolites. Almost all North Atlantic right whales are individually identified, most of the population is sighted each year, and systematic survey effort extends back to 1986. North Atlantic right whales number <500 individuals and are subject to anthropogenic mortality, morbidity and other stressors, and scientific data to inform conservation planning are recognized as important. Here, we describe the use of classification trees as an alternative method of analysing multiple-hormone data sets, building on univariate models that have previously been used to describe hormone profiles of individual North Atlantic right whales of known reproductive state. Our tree correctly classified the age class, sex and reproductive state of 83% of 112 faecal samples from known individual whales. Pregnant females, lactating females and both mature and immature males were classified reliably using our model. Non-reproductive [i.e. 'resting' (not pregnant and not lactating) and immature] females proved the most unreliable to distinguish. There were three individual males that, given their age, would traditionally be considered immature but that our tree classed as mature males, possibly calling for a re-evaluation of their reproductive status. Our analysis reiterates the importance of considering the reproductive state of whales when assessing the relationship between cortisol concentrations and stress. Overall, these results confirm findings from previous univariate statistical analyses, but with a more robust multivariate approach that may prove useful for the multiple-analyte data sets that are increasingly used by conservation physiologists.

Key words: Cetacean, classification tree, *Eubalaena*, faecal sample, non-invasive

Introduction

Increasing industrialization of the ocean has caused concern about potential effects of sublethal stressors on marine wildlife (Pompa *et al.*, 2011; Davidson *et al.*, 2012; Wright, 2012). Assessing sublethal effects on large cetaceans poses particular problems, as whale behaviour is relatively cryptic, their habitat is often logistically difficult to access, and it is impossible to collect plasma samples for traditional physiological studies. Given these constraints, a major advance was

the demonstration that immunoassay of hormone metabolites extracted from faecal samples of free-ranging North Atlantic right whales (*Eubalaena glacialis* Müller 1776, hereafter NARW) provides biologically relevant information on reproductive state and stress responses (Rolland *et al.*, 2005, 2007; Hunt *et al.*, 2006). For example, there is concern that underwater noise from shipping is a stressor for whales (e.g. Wright *et al.*, 2007; Tyack, 2008; Erbe, 2012). Research using faecal endocrine assays, in conjunction with the reduction in noise from shipping after 11 September 2001 as a treatment effect, suggests that these concerns are valid (Rolland *et al.*, 2012).

However, interpretation of faecal hormone data is not simple (Palme *et al.*, 2005; Touma and Palme, 2005; Sheriff *et al.*, 2011; Goymann, 2012). Variables such as sex, diet, season, individual variability, antibody cross-reactivities and even sample mass can influence faecal metabolite concentrations (Touma and Palme, 2005; Hunt *et al.*, 2006; Hayward *et al.*, 2010; Sheriff *et al.*, 2011; Goymann, 2012; Stetz *et al.*, 2013; Kalliokoski *et al.*, 2015). Despite these concerns, recent analyses indicate that faecal hormones are indeed useful for physiological assessment of wildlife and, in some cases, may be superior to plasma-based measures, especially for detection of chronic stress and anthropogenic stress (Sheriff *et al.*, 2011; Dickens and Romero, 2013; Dantzer *et al.*, 2014). Nonetheless, it remains a topic of debate whether faecal hormone metabolites accurately reflect the animal's true physiological state. An additional complication is the lack of a robust published statistical method for combining information from multiple faecal hormones and comparing them directly with independently confirmed physiological state, e.g. as known from individual life-history data (known sex, known reproductive state and age class).

In this regard, the NARW offers an ideal model. Almost all of the (450–500) NARWs are individually identified, 80% of the population has been sighted each year of this study (Hamilton *et al.*, 2007), and systematic survey effort extends back to 1986 (Brown *et al.*, 2007). The NARWs occur seasonally in six well-defined habitats from Florida to the Canadian Maritimes (Kraus and Rolland, 2007), and most individuals are of known age, sex and reproductive state. Calving events are captured by a comprehensive series of aerial surveys of the calving grounds in the southeastern USA (e.g. Keller *et al.*, 2012) and on the northern feeding grounds (e.g. Cole *et al.*, 2013). This comprehensive coverage is coupled with the observation that few juvenile whales enter the photo-identified population that cannot be accounted for by known calving events (Hamilton *et al.*, 2007). This means that the reproductive history was known with certainty for all the females for which faecal samples were collected and identified to that individual.

Since 1999, >375 NARW faecal samples have been collected and analysed for a suite of reproductive and adrenal steroid hormone metabolites (Rolland *et al.*, 2005, 2007; Hunt *et al.*, 2006, 2015). Dietary variation is minimal; NARWs feed on only a few species of copepods, and they actively seek out the C5 stage of *Calanus finmarchicus* (e.g.

Baumgartner and Mate, 2003; Baumgartner *et al.*, 2003) in particular. This minimizes any contribution of dietary variation to faecal hormone concentrations. Furthermore, samples have been collected primarily from a limited geographical region (encompassing the Bay of Fundy, Canada, and the Nova Scotian Shelf), and primarily from July to October, minimizing potential effects of site variation and seasonality. This NARW faecal archive thus provides an excellent sample set with which to test the reliability of faecal hormone metabolites as a proxy measure of physiological condition.

Our goal was to perform a comprehensive test of whether faecal hormone metabolites reflect a whale's true (independently known) physiological state, using a new analytical method for assessing multiple-hormone data sets. To do this, we combined data sets of faecal reproductive steroid hormone metabolites (progestins, androgens and oestrogens) and faecal glucocorticoid (fGC) metabolites (adrenal hormones that increase during physiological stress), using samples that were definitively assigned to an identified individual right whale. A subset of these data (samples from 1999–2004) have been previously analysed for the reproductive steroids (Rolland *et al.*, 2005) and glucocorticoids (Hunt *et al.*, 2006) using multiple, separate univariate approaches; here, we combine those data, and data collected in subsequent years, into a unified multivariate approach.

We combined all four hormone data sets with a novel machine-learning approach to construct an 'evolutionary' variant of a classification tree (Grubinger *et al.*, 2014). Classification and regression trees (CART) are a data-mining (Hastie *et al.*, 2001) technique, producing a decision tree that classifies which reproductive category a sample belongs to, based on hormone concentrations. The tree provides an heuristic model of the data and works by recursive binary splitting. Each split results in two mutually exclusive groups (nodes) that are as homogeneous as possible, based on the response variable, and then each smaller (child) node is further split in a similar manner. Trees have the advantage of being able to handle 'messy' data (e.g. De'ath and Fabricius, 2000), including data that are not multivariate normal, and so are not amenable to analysis using standard multivariate techniques. However, classical formulations of trees rely on relatively simple, forward stepwise searches to determine splits. Although efficient, these can lead to splits that are only locally optimal. For this analysis, we used a new analytical approach, evolutionary trees (Grubinger *et al.*, 2014), that implement an evolutionary algorithm to search for a globally optimal tree. Comparative analyses using benchmark data have demonstrated that evolutionary trees can outperform more classical approaches to generating trees (Grubinger *et al.*, 2014).

Materials and methods

Sample collection

From 1999 to 2011, floating faecal samples were located and collected in the vicinity of feeding right whales. Most

....

2

samples were collected opportunistically during photo-identification surveys of summer habitats and with the assistance of scent detection dogs (Rolland et al., 2005, 2006). A few samples were collected opportunistically during surveys of spring feeding habitats. Faecal material was scooped from the water surface using a 300 μm mesh nylon dip net (Sea-Gear, Inc., Melbourne, FL, USA) attached to an extendable boathook and stored frozen until analysis as previously described by Rolland et al. (2005, 2006).

Whale identification and reproductive state classification

When defecation was observed, photographs were taken of the whale for subsequent photo-identification by comparison with images in the NARW Identification Database (Hamilton et al., 2007; Right Whale Consortium, 2012, rwcatalog.neaq.org). Photographic identification was based on the unique pattern of callosities (i.e. raised, roughened patches of skin) on the whale's head, lips and chin, pigmentation and scars (Kraus et al., 1986). A combination of photo-identification and molecular profiling using DNA extracted from faecal samples (i.e. mitochondrial haplotype, microsatellite profiles) was used to associate faecal samples to known whales using criteria described by Gillett et al. (2010) and Doucette et al. (2012).

Faecal sample storage and processing

Faecal samples were stored at −20°C until the end of the field season, and then transported frozen to our laboratory in Boston (MA, USA), where they were stored at −80°C until analysis. Samples from 2000–2005 were analysed within 6 months of collection; samples from 2006–2011 were archived at −80°C and analysed together in late 2011. Faecal hormone metabolites appear to remain stable for multiple years if samples are kept frozen (Hunt and Wasser, 2003).

Faecal samples were processed and analysed using techniques described by Rolland et al. (2005) and Hunt et al. (2006). Briefly, all samples were freeze-dried and pulverized, the resulting powder was mixed well, and steroids were extracted in 90% methanol using a 10:1 ratio of solvent to faecal mass (e.g. 2.0 ml of 90% methanol added to 0.2 g of dried, well-mixed faecal powder). Samples were vortexed for 30 min, centrifuged for 15 min, and the methanol supernatant (containing hormones) was pipetted to vapour-proof cryovials, diluted in appropriate assay buffers, and assayed within 3 months of extraction. Two separate extracts were produced for each sample and assayed in separate assays, and final results were averaged.

Hormone assays

The hormone assays for glucocorticoids, oestrogens, androgens and progestins have been previously validated for NARW faeces and are described in detail by Rolland et al. (2005) and Hunt et al. (2006). Briefly, the progestin and androgen assays are in-house ^{3}H radioimmunoassays using progesterone antibody CL#425 (Munro laboratory, University of California Davis) and testosterone antibody #250 (Niswender laboratory, University of Colorado), respectively. The oestrogen and glucocorticoid assays are double-antibody ^{125}I radioimmunoassay kits ['total-estrogens' assay no. 140 202 and corticosterone assay no. 02-120 103, both from MP Biomedicals (formerly ICN), Costa Mesa, CA, USA]; the manufacturer's protocols were followed except that the glucocorticoid assay was run at half-volume. This particular glucocorticoid assay uses an antibody that was raised against corticosterone but that also detects mammalian faecal metabolites of cortisol in multiple species, including marine mammals (Wasser et al., 2010).

In all four assays, standards, samples and controls were assayed in duplicate and non-specific binding and zero tubes in quadruplicate, and results averaged. Any sample with >10% coefficient of variation within an assay was re-assayed and the original result discarded. Based on each assay's performance on the tails of the standard curve, cut-offs for acceptable percentage bound were set at 10–90% bound for the glucocorticoid assay, 15–85% bound for the androgen and progestin assays, and 20–80% bound for the oestrogen assay; any samples outside these bounds were re-diluted accordingly and re-assayed. If a sample had >20% coefficient of variation in hormone concentration (in nanograms per gram) across the two separate extracts from that sample, a third extract was produced from dried faecal powder and assayed, the outlying result was discarded, and the other two results were averaged; if no clear outlier was apparent, a fourth extract was produced and assayed, and all four results were averaged. All four assays have inter- and intra-assay variation <10% in our laboratory. For further details and antibody cross-reactivities see Rolland et al. (2005) and Hunt et al. (2006).

Data analysis

For samples from whales of known reproductive state, we tested whether faecal hormone metabolites could be used to ascertain their sex and reproductive state reliably. The interaction between sex and reproductive state was the classifying variable, and levels of four hormone metabolites (androgens, progestins, oestrogens and glucocorticoids) were independent variables. Given that this analysis sought to investigate the reliability of faecal hormone analyses, samples that were too small to run in all four hormone assays were excluded. We also excluded whales suspected a priori to have elevated levels of stress hormones attributable to anthropogenic impacts, e.g. whales entangled in fishing gear or struck by ships. Calves and juveniles of unknown sex were also excluded.

The NARW sex and reproductive state categories were as follows: mature male (MM); immature male (IM); immature female (IF); pregnant female (Preg); lactating female (Lact); and resting (mature, non-pregnant, non-lactating) female (Rest). Life-history data on identified whales (i.e. age, age class, sex and reproductive history) were obtained from the

NARW Identification Database (Right Whale Consortium, 2012). For whales not sighted as calves, an estimated minimal age was used based on the year of first sighting. Classification of age classes followed standard practice with NARW research using the NARW Identification Database (Hamilton *et al.*, 1998): calves, birth to 1 year of age; juveniles, 1–8 years of age; and adults ≥9 years of age or first documented calving for females (if earlier). Trees were produced using the *evtree* library (Grubinger *et al.*, 2014) in R 3.2.0 (R Core Team, 2015).

Results

Sample collection and whale identification

The final data set included 112 samples for which the sex and reproductive state of the NARW that produced the sample could be definitively ascribed. The vast majority of these (104 of 112, 93%) were collected from 1999–2011 in the vicinity of feeding NARWs in the lower Bay of Fundy and Roseway Basin, Canada, where right whales congregate seasonally to feed (July–October). The remaining samples were collected in the spring months (April–June) in right whale habitats in the southern Gulf of Maine. Samples came from 81 identified individual whales, and there were 24 individuals for which hormone data included repeat samples. As there was no *a priori* reason to assume that these should be pseudoreplicates (Hurlbert, 1984), given the purpose of our analyses, these samples were all included in analyses.

Analysis

An initial tree (not shown) comprised seven terminal nodes classifying the samples and included three males identified as 'immature' using the standard classification by age used in the NARW Identification Database (see above). Two of these males were known to be 8 years old at sampling (as they were identified as calves), suggesting that they were likely to be nearly mature or pubescent. The third whale was estimated to be at least 6 years old; however, he was not identified in his calving year, and therefore may have been >9 years old when the sample was collected. Faecal androgen levels for these three individuals were 6459, 7803 and 6250 ng/g, all substantially higher than the mean + 2 SEM for androgens previously recorded for juvenile NARW (5558.2 ng/g, as 4422 ± 568.1 ng/g is mean + 1 SEM; Rolland *et al.*, 2005). Based upon their elevated faecal androgen levels (a direct correlate of testicular activity that may be a more accurate indicator of sexual maturity than age; see, for example, Beehner *et al.*, 2009), these three males were reclassified as 'mature' and the analysis was re-run. Levels of reproductive steroids, as classified using the Database's standards (Fig. 1; note that progesterone and oestrogen levels are $\log_{10}$ transformed to improve the readability of those plots), varied with sex and reproductive state as has been previously described using a subset of the samples analysed here (Rolland *et al.*, 2005).

The tree (Fig. 2) had an overall misclassification of 17.0%. Pregnant females, lactating females and mature males were categorized with excellent reliability by nodes in the tree (Table 1). Most immature males were categorized correctly, but the tree had the greatest difficulty distinguishing between immature and resting females (Fig. 2 and Table 1). 'Nodes' in the tree are where the tree splits into two separate branches. 'Terminal nodes' are where the tree branches no further. The low rate of successful classification of resting females was because several (nine) were classified in a node identified by the algorithm as being immature females (terminal node 8 in Fig. 2; all nodes hereafter refer to Fig. 2), but this node really represented a mix of immature and resting females, which are both in a non-reproductive state. In an earlier analysis, we also found that immature and resting females were indistinguishable using faecal reproductive hormone data (Rolland *et al.*, 2005). As there were only six separate sex–reproductive state categories in the analysis, but seven terminal nodes in the tree, one category (immature females) was represented in two nodes, 4 and 8 (Fig. 2). The misclassification rate for individuals other than immature and resting females was 10.5%.

Highly elevated levels of faecal progestins (≥11 500 ng/g) separated pregnant females from all other individuals (terminal node 13). Next, levels of fGC >16.3 ng/g separated resting females and some immature females (with lower levels of fGC) from other whales (node 2). The 'low-fGC' node then split into two terminal nodes; resting females had higher levels of progestins (terminal node 5) than a terminal 'immature female' node that included three resting females (terminal node 4). The 'higher fGC' node split (node 6), based on oestrogen levels, into a node of immature animals (node 7) with lower oestrogen and a node combining lactating females and mature males (node 10), which split further based on androgen levels (terminal nodes 11 and 12). The 'lower oestrogen' node of immature whales split (node 7) based on androgen levels into terminal nodes of immature males (terminal node 9, higher androgens) and a terminal 'immature female' node that included five resting females (terminal node 8).

Glucocorticoid levels in samples categorized correctly (i.e. to correct sex and reproductive class) by the tree were different (Kruskal–Wallis test, for Testosterone, $\chi^2_5 = 60.7$, $P < 0.001$; Fig. 1).

Discussion

The aim of any tree algorithm (evolutionary trees, CARTs or others) is to balance predictive accuracy and complexity, while producing a model that can be easy to interpret (Grubinger *et al.*, 2014). In this instance, we have a tree that has a reasonable level of misclassification (five out of every six samples were classified correctly) and separates out most sex, age and reproductive classes in an intuitive manner. We did not aim, in this analysis, to develop a definitive descriptor of the hormone metabolite levels that classify right whales into different classes of reproductive states (see Rolland *et al.*, 2005; Hunt
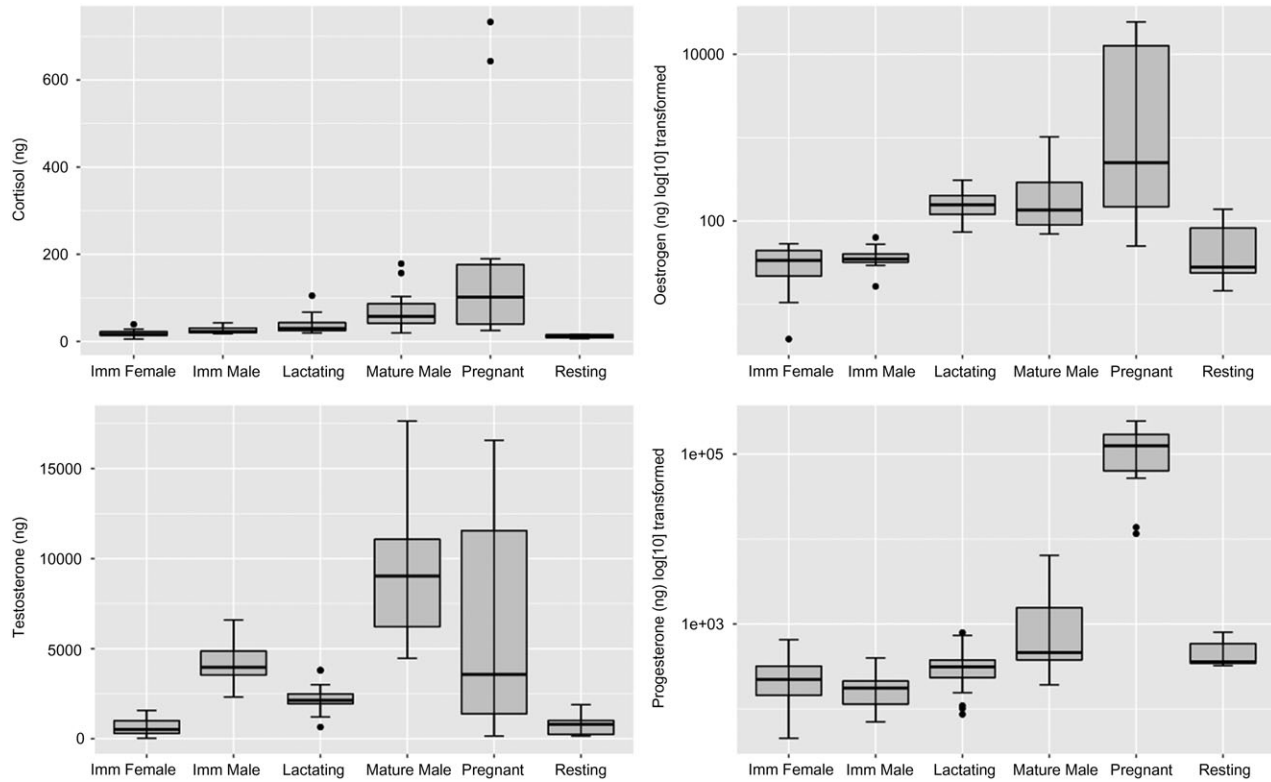
Figure 1: Boxplots of hormone metabolites extracted from faecal samples of North Atlantic right whales (*Eubalaena glacialis*). Plots were produced using the *ggplot2* library (Wickham, 2009) of R 3.3.1 running under RStudio. Upper and lower limits of the box are 25% and 75% quantiles; box mid-lines are medians. Upper and lower whiskers show values to 1.5 times the distance between the 25 and 75% quartiles. Dots indicate outliers. Oestrogen and progesterone values are log_{10} transformed.

et al., 2006 for baseline faecal hormone ranges in the NARW) but to verify the previous, univariate approaches using an analytical technique that combines all data in one analysis. To our knowledge, this is the first use of this type of technique in analysing faecal hormones of any free-ranging wildlife.

Given these results, we are confident that faecal hormones reliably reflect predicted physiological states in this species. Multiple studies now indicate that this is the case in many other mammals as well, despite the noise introduced by the myriad other variables that can potentially affect faecal hormone concentrations (see Introduction). For example, in dugongs (*Dugong dugon*), pregnancy can be reliably detected from faecal progestins, the progesterone concentrations derived from serum and faecal samples were highly correlated, and faecal androgens were reliable indicators of sexual maturation and reproductive activity in males (Burgess *et al.*, 2012a,b). Likewise, in killer whales (*Orcinus orca*), faecal progestins and androgens predicted reproductive state; and glucocorticoids and thyroid hormones have been successfully used to distinguish between the variable effects of boat traffic and prey availability (Ayres *et al.*, 2012).

Two recent reviews have concluded that faecal hormone analysis may even be superior to plasma-based measures,

particularly for assessment of chronic stress and responses to anthropogenic stressors (Dickens and Romero, 2013; Dantzer *et al.*, 2014). Goymann (2012) and others have pointed out that possible effects of diet, season, temperature and other influences on faecal hormone data should not be overlooked. We suggest that the best way to address such concerns is to validate faecal hormone analysis in populations that have known individuals, enabling comparison of faecal hormone profiles directly with individual state (sex, age class and reproductive state).

North Atlantic right whales may represent an ideal case for faecal hormone analysis because several of the potentially confounding variables discussed by Goymann (2012) are naturally minimized. North Atlantic right whales have very little dietary variation (Baumgartner *et al.*, 2007) and tend to feed within a relatively narrow range of water temperatures (minimizing the dramatic seasonal changes in metabolic rate discussed by Goymann, 2012). Furthermore, sample degradation issues, such as those discussed by Stetz *et al.* (2013), are minimal because faecal samples in marine studies are typically collected within minutes of defecation, unlike terrestrial studies where samples may be collected days to weeks
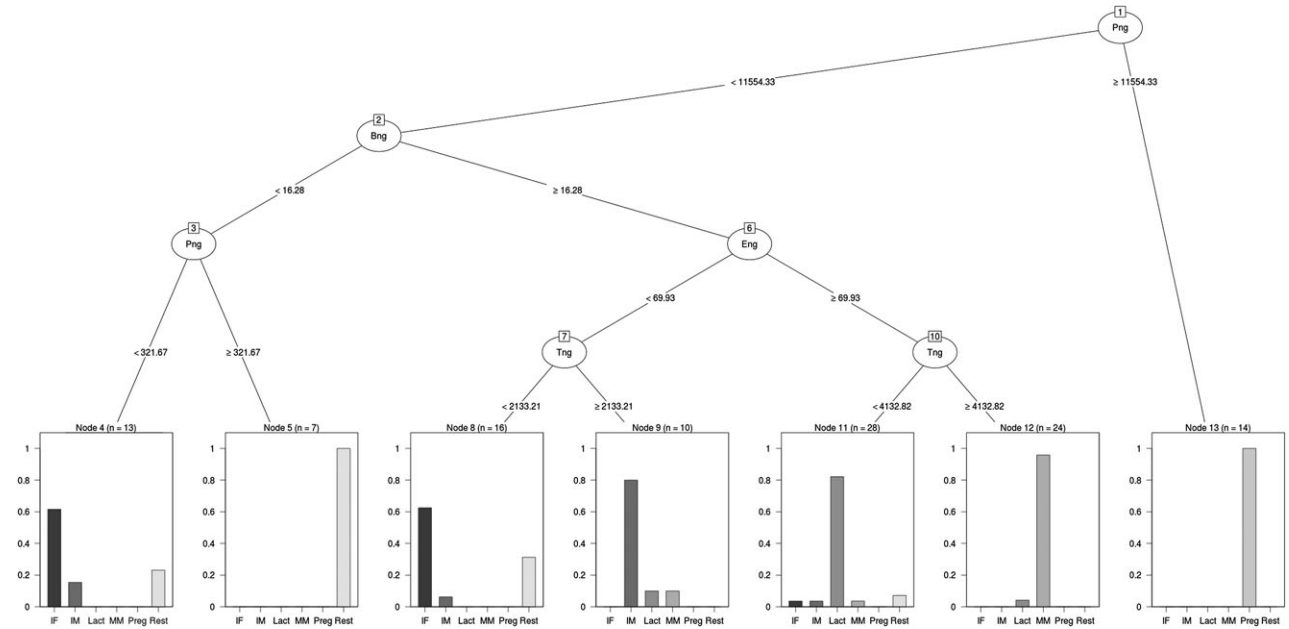
Figure 2: Tree of known reproductive states of North Atlantic right whales (*Eubalaena glacialis*) constructed using the *evtree* library (Grubinger *et al.*, 2014) of R 3.3.1 running under RStudio. The NARW sex and reproductive state categories were as follows: mature male (MM); immature male (IM); immature female (IF); pregnant female (Preg); lactating female (Lact); and resting (mature, non-pregnant, non-lactating) female (Rest). The variables used for splitting (all values in nanograms of immunoreactive hormone per gram of dried faecal powder) are oestrogen (Eng); Progesterone (Png); Testosterone (Tng); and Cortisol (Bng). Names in the ovals at each node of the tree indicate the variable driving the split at that node. Numbers in the small boxes on the ovals at each node indicate the node's number. Values on the lines joining nodes (e.g. >11 554.33 on the line joining nodes 1 and 13; <11 554.33 on the line joining nodes 1 and 2) indicate the value of the variable (in nanograms of immunoreactive hormone per gram of dried faecal powder) in each direction of the split. Bar graphs at the terminal nodes of the tree show the proportion of that node that is composed of each of the known reproductive states.

Table 1: Accuracy of the classification tree of North Atlantic right whale age and sex classes from faecal hormones

Reproductive class	Sample size	Correctly classified	Incorrectly classified	Percentage correctly classified
Mature male	25	23	2	92.0
Immature male	12	8	4	66.7
Pregnant female	14	14	0	100.0
Lactating female	25	23	3	92.0
Resting female	17	7	10	41.2
Immature female	19	18	1	94.7

after excretion. Likewise, the dugong and killer whale studies discussed above (Ayres *et al.*, 2012; Burgess *et al.*, 2012a,b) were also characterized by a relatively consistent diet, rapid sample collection and minimal temperature variation. Certain species may thus be more amenable to faecal hormone analysis than others.

Faecal hormone analyses also may offer a method to assess the onset of maturity in male NARWs. Determining the age at sexual maturity of baleen whales is challenging, especially for males. Samples obtained from past commercial whaling operations allowed for reliable estimation for females, comparing age (from ear plug laminations) with ovarian state, and likewise for males by relating testicular development with age from laminations (e.g. Best and Lockyer, 2002). Using non-invasive techniques, age at first calving for females can be determined observationally, from photo-identification catalogues that have sufficient coverage of their study populations to ensure that a first calving event is unlikely to be missed (e.g. Clapham and Mayo, 1990; Kraus *et al.*, 2007). Onset of the age of sexual maturity for male baleen whales using these observational techniques is inherently more difficult because they lack an attendant calf.

Also, the onset of sexual maturity is one of the more labile life-history parameters in mammals (e.g. Gaillard *et al.*, 2000). At present, the NARW Identification Database's approach to determining whether a whale is mature is age based (Hamilton *et al.*, 2007). The age of maturation in males is estimated at 9 years based on the mean age of female maturity (Kraus *et al.*, 2007). In our original analyses, three 'immature' male whales could be classified into the 'mature' node by their androgen levels; two of these were known with certainty to be 8 years old at the time of sampling. Additionally, a 10-year-old male was classified in the final tree as 'immature' (node 9 in Fig. 2). This whale may have been pubertal, as it had a faecal androgen level of 6584 ng/g, substantially higher than the baseline for immature males (Rolland *et al.*, 2005). Although these sample sizes are admittedly small, our data indicate that male NARWs may reach sexual maturity between the ages of 8 and 10 years.

Our tree-based method reliably detects reproductively active adult females (pregnant and lactating) and adult males. However, with a misclassification rate of one in six, our method does not clearly determine the sex and reproductive state of all whales accurately enough that samples from unknown whales in other reproductive categories can be categorized with complete certainty. Considering this caveat, there are potential conservation benefits (Rolland *et al.*, 2007; Hunt *et al.*, 2013) derived from further analyses of faecal hormones. Baleen whales can show patterns of movement that are differentiated by sex, which has implications for estimating population size (e.g. Brown *et al.*, 1995; Valsecchi *et al.*, 2010). Migration patterns also vary by reproductive class (e.g. Dawbin, 1966), with consequences for the likely susceptibility of different classes (e.g. pregnant females) to anthropogenic stressors. Although sex determination of large whales is generally made by genetic analysis of skin samples obtained by remote biopsy, photographs of an identified individual's genital slit, or by observations of identified individuals with a calf (Hamilton *et al.*, 2007), faecal hormone analyses provide a secondary check. Furthermore, the ability to identify pregnancies and potentially assess sexual maturity could result in more accurate estimates of the proportion of breeding individuals in a population.

Two recent reviews have concluded that fGC analysis may be superior to plasma-based measures, particularly for assessment of chronic stress and responses to anthropogenic stressors (Dickens and Romero, 2013; Dantzer *et al.*, 2014). In NARWs, fGCs have proved useful for identifying exposure to chronic stressors (Rolland *et al.*, 2012). In our analysis, glucocorticoid levels were also confirmed to vary with reproductive state, as previously shown (Hunt *et al.*, 2006; Rolland *et al.*, 2007), with resting and immature females having the lowest levels of fGCs (as shown by the second split in the tree, node 2 in Fig. 2). This can complicate the identification of drivers of observed elevations of glucocorticoids, especially if the sample was collected from an unknown individual. It is clear from our data that fGCs can be elevated in healthy animals of certain reproductive states, namely, pregnant and lactating females and mature males (Figs 1 and 2;

also see Hunt *et al.*, 2006). Breeding activity and pregnancy are known to cause elevations in glucocorticoids in other species as well (e.g. Dantzer *et al.*, 2010). Thus, studies focused on determining stress responses to anthropogenic factors should use reproductive hormone data along with glucocorticoid data to control for natural variations in stress hormones with reproductive state.

Our 13 year data set indicates that hormone metabolites from right whale faecal samples are reliable indicators of the physiological state of individual whales. By taking a machine-learning approach to our data analysis, we avoided any possible issues of 'researcher degrees of freedom' (i.e. circular reasoning; Simmons *et al.*, 2011) influencing our results. We believe the approach presented here provides a complementary analytical option to the univariate, hormone-by-hormone analyses traditionally used in faecal hormone studies. The results of the tree also offer a way to classify faecal samples by age and reproductive category when those samples are not linked to a known individual whale, offering a new way to assess the demographic structure of whale populations (see also Labrada-Martagón *et al.*, 2014). These can then be further checked against the baseline data already derived using standard techniques (Rolland *et al.*, 2005; Hunt *et al.*, 2006).

Acknowledgements

Funding

References

Ayres KL, Booth RK, Hempelmann JA, Koski KL, Emmons CK, Baird RW, Balcomb-Bartok K, Hanson MB, Ford MJ, Wasser SK (2012) Distinguishing the impacts of inadequate prey and vessel traffic on an endangered killer whale (*Orcinus orca*) population. *PLoS One* 7: e36842.

Baumgartner MF, Mate BR (2003) Summertime foraging ecology of North Atlantic right whales. *Mar Ecol Prog Ser* 264: 123–135.

Baumgartner MF, Cole TVN, Campbell RG, Teegarden GJ, Durbin EG (2003) Associations between North Atlantic right whales and their prey, *Calanus finmarchicus*, over diel and tidal time scales. *Mar Ecol Prog Ser* 264: 155–166.

Baumgartner MF, Mayo CA, Kenney RD (2007) Enormous carnivores, microscopic food and a restaurant that's hard to find. In SD Kraus, RM Rolland, eds, *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA, pp 138–171.

Beehner JC, Gesquiere L, Seyfarth RM, Cheney DL, Alberts SC, Altmann J (2009) Testosterone related to age and life-history stages in male baboons and geladas. *Horm Behav* 56: 472–480.

Best PB, Lockyer CH (2002) Reproduction, growth and migrations of sei whales (*Balaenoptera borealis*) off the west coast of South Africea in the 1960s. *S Afr J Mar Sci* 24: 111–133.

Brown MR, Corkeron PJ, Hale PT, Schultz KW, Bryden MM (1995) Evidence for a sex-segregated migration in humpback whales, *Megaptera novaeangliae*. *Proc Biol Sci* 259: 229–234.

Brown MW, Kraus SD, Slay CK Garrison LP (2007) Surveying for science, discovery, and management. In SD Kraus, RM Rolland, eds, *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA, pp 105–137.

Burgess EA, Lanyon JM, Brown JL, Blyde D, Keeley T (2012a) Diagnosing pregnancy in free-ranging dugongs using fecal progesterone metabolite concentrations and body morphometrics: a population application. *Gen Comp Endocrinol* 177: 82–92.

Burgess EA, Lanyon JM, Keeley T (2012b) Testosterone and tusks: maturation and seasonal reproductive patterns of live, free-ranging male dugongs (*Dugong dugon*) in a subtropical population. *Reproduction* 143: 683–697.

Clapham PJ, Mayo CA (1990) Reproduction of humpback whales (*Megaptera novaeangliae*) observed in the Gulf of Maine. *Rep Int Whaling Comm Spec Iss* 12: 171–175.

Cole TVN, Hamilton P, Henry AG, Duley P, Pace RM III, White BN, Frasier T (2013) Evidence of a North Atlantic right whale *Eubalaena glacialis* mating ground. *Endang Species Res* 21: 55–64.

Dantzer B, McAdam AG, Palme R, Fletcher QE, Boutin S, Humphries MM, Boonstra R (2010) Fecal cortisol metabolite levels in free-ranging North American red squirrels: assay validation and the effects of reproductive condition. *Gen Comp Endocrinol* 167: 279–286.

Dantzer B, Fletcher QE, Boonstra R, Sheriff MJ (2014) Measures of physiologic stress: a transparent or opaque window into the status, management and conservation of species? *Conserv Physiol* 2: cou023; doi:10.1093/conphys/cou023.

Davidson AD, Boyer AG, Kim H, Pompa-Mansilla S, Hamilton MJ, Costa DP, Ceballos G, Brown JH (2012) Drivers and hotspots of extinction risk in marine mammals. *Proc Natl Acad Sci USA* 109: 3395–3400.

Dawbin WH (1966) The seasonal migratory cycle of humpback whales. In Norris KS, ed, *Whales, Dolphins and Porpoises*. University of California Press, Berkley and Los Angeles, CA, pp 145–170.

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.

Dickens MJ, Romero LM (2013) A consensus endocrine profile for chronically stressed wild animals does not exist. *Gen Comp Endocrinol* 191: 177–189.

Doucette GJ, Mikulski CM, King KL, Roth PB, Wang Z, Leandro LF, DeGrasse SL, White KD, De Biase D, Gillett RM *et al.* (2012) Endangered North Atlantic right whales (*Eubalaena glacialis*) experience repeated, concurrent exposure to multiple environmental neurotoxins produced by marine algae. *Environ Res* 112: 67–76.

Erbe C (2012) Effects of underwater noise on marine mammals. *Adv Exp Med Biol* 730: 17–22.

Gaillard J-M, Festa-Bianchet M, Yoccoz NG, Loison A, Toïgo C (2000) Temporal variation in fitness components and population dynamics of large herbivores. *Annu Rev Ecol Syst* 31: 367–393.

Gillett RM, Frasier TR, Rolland RM, White BN (2010) Molecular identification of individual North Atlantic right whales (*Eubalaena glacialis*) using free-floating feces. *Mar Mammal Sci* 26: 917–936.

Goymann W (2012) On the use of non-invasive hormone research in uncontrolled, natural environments: the problem with sex, diet, metabolic rate and the individual. *Methods Ecol Evol* 3:757–765.

Grubinger T, Zeileis A, Pfeiffer K (2014) Evtree: evolutionary learning of globally optimal classification and regression trees in R. *J Stat Softw* 61: 1–29. http://www.jstatsoft.org/v61/i01/.

Hamilton PK, Knowlton AR, Marx MK, Kraus SD (1998) Age structure and longevity in North Atlantic right whales (*Eubalaena glacialis*) and their relation to reproduction. *Mar Ecol Prog Ser* 171: 285–292.

Hamilton PK, Knowlton AR, Marx MK (2007) Right whales tell their own stories: the photo-identification catalog. In SD Kraus, RM Rolland, eds, *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA, pp 75–104.

Hastie T, Tibshirani R, Friedman J (2001) *The Elements of Statistical Learning*. Springer, New York, NY.

Hayward LS, Booth RK, Wasser SK (2010) Eliminating the artificial effect of sample mass on avian fecal hormone metabolite concentration. *Gen Comp Endocrinol* 169: 117–122.

Hunt KE, Wasser SK (2003) Effect of long-term preservation methods on fecal glucocorticoid concentrations of grizzly bear and African elephant. *Physiol Biochem Zool* 76: 918–928.

Hunt KE, Rolland RM, Kraus SD, Wasser SK (2006) Analysis of fecal glucocorticoids in the North Atlantic right whale (*Eubalaena glacialis*). *Gen Comp Endocrinol* 148: 260–272.

Hunt KE, Moore MJ, Rolland RM, Kellar NM, Hall AJ, Kershaw J, Raverty SA, Davis CE, Yeates LC, Fauquier DA *et al.* (2013) Overcoming the challenges of studying conservation physiology in large whales: a review of available methods. *Conserv Physiol* 1: cot006; doi:10.1093/conphys/cot006.

Hunt KE, Rolland RM, Kraus SD (2015) Conservation physiology of an uncatchable animal: the North Atlantic right whale (*Eubalaena glacialis*). *Integr Comp Biol* 55: 577–586.

Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* 54: 187–211

Kalliokoski O, Teilmann AC, Abelson KSP, Hau J (2015) The distorting effect of varying diets on fecal glucocorticoid measurements as indicators of stress: a cautionary demonstration using laboratory mice. *Gen Comp Endocrinol* 211: 147–153.

Keller CA, Garrison L, Baumstark R, Ward-Geiger LI, Hines E (2012) Application of a habitat model to define calving habitat of the North Atlantic right whale in the southeastern United States. *Endang Species Res* 18: 73–87.

Kraus SD, Moore KE, Price CE, Crone MJ, Watkins WA, Winn HE, Prescott JH (1986) The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). *Rep Int Whal Comm Spec Iss* 10: 145–151.

Kraus SD, Pace RM III, Frasier TR (2007) High investment, low return: the strange case of reproduction in *Eubalaena glacialis*. In SD Kraus, RM Rolland, eds, *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA, pp 172–199.

Kraus SD, Rolland RM (2007) Right whales in the urban ocean. In SD Kraus, RM Rolland, eds, *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA, pp 1–38.

Labrada-Martagón V, Zenteno-Savín T, Mangel M (2014) Linking physiological approaches to marine vertebrate conservation: using sex steroid hormone determinations in demographic assessments. *Conserv Physiol* 2: cot035; doi:10.1093/conphys/cot035.

Palme R, Rettenbacher S, Touma C, el-Bahr SM, Möstl E (2005) Stress hormones in mammals and birds: comparative aspects regarding metabolism, excretion, and noninvasive measurement in fecal samples. *Ann NY Acad Sci* 1040: 162–171.

Pompa S, Ehrlich PR, Ceballos G (2011) Global distribution and conservation of marine mammals. *Proc Natl Acad Sci USA* 108: 13600–13605.

R Core Team (2015) *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Right Whale Consortium (2012). North Atlantic Right Whale Consortium Identification and Sightings Databases 31 December 2012 (New England Aquarium, Boston, MA, USA).

Rolland RM, Hunt KE, Kraus SD, Wasser SK (2005) Assessing reproductive status of right whales (*Eubalaena glacialis*) using fecal hormone metabolites. *Gen Comp Endocrinol* 142: 308–317.

Rolland RM, Hamilton PK, Kraus SD, Davenport B, Bower RM, Wasser SK (2006) Faecal sampling using detection dogs to study reproduction and health in North Atlantic right whales (*Eubalaena glacialis*). *J Cetacean Res Manage* 8: 121–125.

Rolland RM, Hunt KE, Doucette GJ, Rickard LG, Wasser SK (2007) The inner whale: hormones, biotoxins and parasites. In SD Kraus, RM Rolland, eds, *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA, pp 232–272.

Rolland RM, Parks SE, Hunt KE, Castellote M, Corkeron PJ, Nowacek DP, Wasser SK, Kraus SD (2012) Evidence that ship noise increases stress in right whales. *Proc Biol Sci* 279: 2363–2368.

Sheriff MJ, Dantzer B, Delehanty B, Pame R, Boonstra R (2011) Measuring stress in wildlife: techniques for quantifying glucocorticoids. *Oecologia* 166: 869–887.

Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22: 1359–1366.

Stetz J, Hunt K, Kendall KC, Wasser SK (2013) Effects of exposure, diet, and thermoregulation on fecal glucocorticoid measures in wild bears. *PLoS One* 8: e55967.

Touma C, Palme R (2005) Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. *Ann NY Acad Sci* 1046: 54–74.

Tyack PL (2008) Implications for marine mammals of large-scale changes in the marine acoustic environment. *J Mammal* 89: 549–558.

Valsecchi E, Corkeron PJ, Galli P, Sherwin W, Bertorelle G (2010) Genetic evidence for sex-specific migratory behaviour in western South Pacific humpback whales. *Mar Ecol Prog Ser* 398: 275–286.

Wasser SK, Cristobal-Azkarate JA, Booth RN, Hayward L, Hunt K, Ayres K, Vynne C, Gobush K, Canales-Espinosa D, Rodriguez-Luna E (2010) Non-invasive measurement of thyroid hormone in feces of a diverse array of avian and mammalian species. *Gen Comp Endocrinol* 168: 1–7.

Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, NY.

Wright AJ (2012) Noise-related stress and cumulative impact assessment. *Adv Exp Med Biol* 730: 541–543.

Wright AJ, Soto NA, Baldwin AL, Bateson M, Beale CM, Clark C, Deak T, Edwards EF, Fernandez A, Godinho A *et al.* (2007) Do marine mammals experience stress related to anthropogenic noise? *Int J Comp Psychol* 20: 274–316.