



Characterizing sex ratios of sea turtle populations: A Bayesian mixture modeling approach applied to juvenile loggerheads (*Caretta caretta*)

Kyle W. Shertzer^{a,*}, Larisa Avens^a, Joanne Braun McNeill^a, April Goodman Hall^a, Craig A. Harms^b

^a National Marine Fisheries Service, Southeast Fisheries Science Center, Beaufort, NC, United States

^b North Carolina State University, College of Veterinary Medicine, Center for Marine Sciences and Technology, 303 College Circle, Morehead City, NC, United States

ARTICLE INFO

Keywords:

Bayesian
Endangered
Loggerhead
Mixture models
Sea turtles
Sex ratio

ABSTRACT

Sex ratio is fundamental to the demography of animal populations. For sea turtles, the operational sex ratio (i.e., that of breeding adults) can be inferred from juvenile sex ratio. In this study, we characterize sex ratio of $n = 1401$ juvenile loggerhead sea turtles (*Caretta caretta*) sampled from foraging grounds in North Carolina, USA. For a subset ($n = 170$), sex was confirmed by laparoscopy, and for the remaining individuals, sex was classified probabilistically using a mixture regression model. The model predicted plasma testosterone concentration as a function of sex, which was treated as a latent variable, and of other potential covariates, namely water temperature and carapace length. Furthermore, it quantified uncertainty in the sex classification of each individual. For the full data set, the predicted sex ratio was 81% female. Using the subset of turtles with known sex and a leave-one-out validation approach, we found the model to have classification accuracy of 94%. We further used that subset to examine how many laparoscopies are sufficient for accurate classification of turtles with unknown sex, in an attempt to provide guidance for other studies. We found diminishing returns as sample size increased, and recommend 100–140 turtles as a sufficient range.

1. Introduction

All sea turtle species are categorized on the IUCN Red List of Threatened species as vulnerable, endangered, critically endangered, or data deficient if accurate assessment is not possible (www.iucnredlist.org). Demographic data are essential for developing models to assess sea turtle population status, as well as to predict the potential conservation benefits of management actions (Heppell et al., 2003a, 2003b). While sex ratio is a fundamental component in the demography of animal populations overall, in the case of sea turtles it is of particular interest because hatchling sex is influenced by egg incubation temperature, raising the prospect that sex ratio will be – or is – affected by climatic warming (Davenport, 1997; Hays et al., 2014; Laloë et al., 2014; Braun McNeill et al., 2016). If so, estimating and monitoring sex ratios of sea turtle populations will be increasingly important for understanding their population ecology and, ultimately, for conservation (Hamann et al., 2010; Rees et al., 2016).

To understand population growth rates, the operational sex ratio (i.e., the sex ratio of breeding adults) is of primary interest. However, focusing on the juvenile stage of sea turtle populations has been recommended for inferring operational sex ratios (Braun McNeill et al., 2016; Wibbels, 2003), because 1) doing so avoids potential sex-specific

breeding behaviors that might bias sampling of adults, 2) hatchling sex ratio might not propagate to adults if male and female hatchlings experience different mortality rates, and 3) due to a long developmental period, juvenile populations comprise numerous cohorts of hatchlings produced at various nesting beaches. As juvenile sea turtles lack external sex-linked characteristics, various methods have been attempted to determine sex for this life stage, such as genetic and endocrinological analyses, as well as laparoscopic examination of gonads (henceforth, laparoscopy) (Wibbels et al., 2000). However, given the suite of available techniques, the approach considered most effective is to measure circulatory plasma testosterone concentrations followed by laparoscopy for a subset of turtles to assign sex-specific ranges of testosterone (Wibbels et al., 2000; Wibbels, 2003). Once these ranges are established, they can be used to assign sex to additional individuals based solely on testosterone concentrations, and ultimately to calculate an overall population sex ratio (Wibbels et al., 2000; Wibbels, 2003).

While this stepwise methodology has been applied frequently and has yielded valuable data (Wibbels et al., 2000), some challenges remain. First, a number of studies report the inability to fully classify turtles based on testosterone concentrations either because female and male ranges overlap, or because there is a substantial gap between the highest known female and lowest known male values. This proportion

* Corresponding author.

E-mail address: Kyle.Shertzer@noaa.gov (K.W. Shertzer).

of “unknown” turtles varies among studies and is often low (0–3%: e.g., Braun McNeill et al., 2016; Geis et al., 2005; León and Diez, 1999; Sanchez, 2013; Wibbels, 1988). However, it can be as high as 5–8% (e.g., Allen et al., 2015; Arendt et al., 2012; Blanvillain et al., 2008; Bolten et al., 1992; Wibbels et al., 1987), and in some cases up to 11–13% (e.g., Geis et al., 2003; Witzell et al., 2005), potentially influencing the ability to calculate population sex ratio accurately. In addition, ambient water temperature can influence testosterone concentrations and, consequently, the ability to assign sex to juvenile sea turtles using this method (Braun McNeill et al., 2007; Hawkes et al., 2013). As a result, more information is needed to determine possible effects of water temperature on testosterone and guide the timing of sampling periods to ensure accurate results (Braun McNeill et al., 2007). Finally, the laparoscopic validation necessary to establish sex-specific testosterone ranges is challenging, as it requires specialized expertise, equipment to conduct the surgeries, and facilities for post-operative observations (Wibbels et al., 2000). Thus, there exists a need for guidelines concerning the number of laparoscopies required to achieve suitable predictive power.

Incorporation of partially-validated testosterone data into a *mixture model* may provide an avenue to address these issues and refine existing approaches for estimating sea turtle population sex ratios. The theory and application of mixture models have received much attention in the fields of statistics (McLachlan and Peel, 2000; Gelman et al., 2013) and machine learning, where the approach is referred to as supervised or unsupervised classification (Hastie et al., 2009). In marine systems, mixture models have proven useful in such applications as analyzing animal movement (Patterson et al., 2009), pinniped demography (Sweeney et al., 2015), fishery discard mortality (Benoit et al., 2012), somatic growth (Shertzer et al., 2017), and stock identification (Millar, 1987; Pella and Masuda, 2001) including stocks of sea turtles (Bolten et al., 1998; Bolker et al., 2007). In particular, a recent study of sex ratio in green sea turtles (*Chelonia mydas*) contained three individuals of unknown sex, and a mixture model was used to classify them as female (Allen et al., 2015).

The goals of this paper are threefold. First, we further test the utility of mixture models for classifying juvenile sea turtle sex using the largest available published testosterone dataset, which contains ~1500 juvenile loggerhead sea turtles (*Caretta caretta*) from which samples were collected over a 14-year period (Braun McNeill et al., 2016). The application considers model selection criteria to inform the choice of model structure and then classifies individual observations (turtles) into their most likely components (male or female). We further quantify uncertainty in these classifications, allowing for one to judge the strength of inference and for the propagation of error to any subsequent analyses. Second, we evaluate the potential influence of covariates such as water temperature on testosterone levels. Third, we examine model performance (classification accuracy) for various proportions of turtles in the data set with known sex, with the intent to provide guidance on how many turtles to examine laparoscopically for establishing sufficient testosterone reference data.

2. Materials and methods

2.1. Data collection and sexing technique

During 1997–2010 from May through November, juvenile loggerhead turtles were sampled from foraging grounds after being captured in pound nets or long-haul seines fished commercially in Core and Pamlico Sounds, North Carolina, USA (Fig. 1). Pound nets are a stationary gear that passively capture targeted fish (Higgins and Pearson, 1928); long-haul seines (here, typically 1 km × 2 m) are pulled between two boats to actively encircle and concentrate the catch (Guthrie et al., 1973). When sea turtles are incidentally captured, the open nature of the nets allows access for sampling. At each encounter, the turtles were brought on board small fishing or research vessels to tag and collect

morphometric data and blood samples. In addition, surface water temperature was recorded to the nearest 0.5 °C using thermometers calibrated to ensure accuracy throughout the study period.

To identify individual turtles, we tagged both rear flippers with Inconel metal alloy size 681 self-piercing tags (National Brand and Tag Company, Newport, Kentucky, USA) and injected a Passive Integrated Transponder (PIT) tag subcutaneously (Destron-Fearing Corp., South St. Paul, Minnesota, USA, 125 or 134 kHz). PIT tags were injected approximately 1 cm anterior to the second most proximal scale of the trailing margin of the left or right front flipper, or into the triceps superficialis muscle of the left front flipper. We measured standard straight-line carapace length (cm) using large calipers, and to prevent use of adults in this study (and avoid bias associated with various sex-specific behaviors of the adult population; Wibbels et al., 1987; Wibbels et al., 1991), we excluded from the analysis turtles with carapace length > 75 cm, as this is the minimum adult size observed for this population of loggerheads (Avens et al., 2015).

For endocrine analysis, we collected 5 ml of blood within 30 min of capture from the dorsocervical sinus of the turtle using a sterile syringe with a 3.81 cm, 20 gauge needle, (Owens and Ruiz, 1980), and we immediately transferred the sample to a sterile lithium heparin or sodium heparin tube which was stored on ice for a maximum of 5 h (i.e., for the rest of the field day). In the laboratory, we centrifuged blood samples for 6–10 min, pipetted 2 ml samples of plasma into cryogenic vials, and stored the plasma samples at –80 °C until analysis. These samples were subsequently processed after each field season using a testosterone radioimmunoassay technique (detailed in Owens et al., 1978; Wibbels et al., 1987; Owens, 1997) that was consistent throughout the course of the study and was previously validated for loggerhead plasma (Owens, 1997). For each assay, multiple aliquots of a loggerhead serum pool were used as a control. The intra-assay CV was 11.4% ± 0.95 (mean ± standard error), and the inter-assay CV was 17.9% ± 3.6. Seventy-two assays were run to process the samples, with a mean extraction efficiency of 87.8% ± 0.17, which was accounted for in the reported testosterone concentrations. The resulting data (after some culling, see Section 2.7), which we refer to as the *full data set*, contained information on $n = 1401$ juvenile loggerhead turtles.

A subset of turtles were selected opportunistically for laparoscopy as previously described (Wood et al., 1983; Wibbels, 1999), and modified in later years to incorporate short-acting general anesthesia with propofol (5 mg/kg IV) in addition to a local anesthetic (lidocaine ≤ 2 mg/kg), short-term postoperative analgesia (ketoprofen 2 mg/kg IM), lateral recumbency (vs. head-down) in a custom-made restraint device, and closure with monofilament nominally absorbable sutures (poligle-caprone 25) with lower tissue reactivity (Maclean et al., 2008). This procedure was conducted at the NOAA Beaufort Laboratory in Beaufort, North Carolina. It verified the sex of $n = 170$ turtles, of which 121 were female and 49 were male, and we refer to this subset as the *training data set* (Supplement A). Thus our data contained 1401 turtles with sex known for 170 individuals and unknown for 1231, and potential covariate information on testosterone concentration, carapace length, and water temperature at the time blood samples were collected.

2.2. Mixture model framework

We fit the natural log of plasma hormone (testosterone) concentrations (H) using mixture regression models (Hamel et al., 2017) implemented in a Bayesian framework (Gelman et al., 2013). Here, the observed log concentrations ($h = h_1, \dots, h_n$) were considered to be generated by a mixture of two components, males ($s = 1$) and females ($s = 2$). We modeled the expected log testosterone concentration (\bar{H}_s) of each sex as a function (f_s) of carapace length (L) and water temperature (T),

$$\bar{H}_s = f_s(L, T | \theta_s) \quad (1)$$

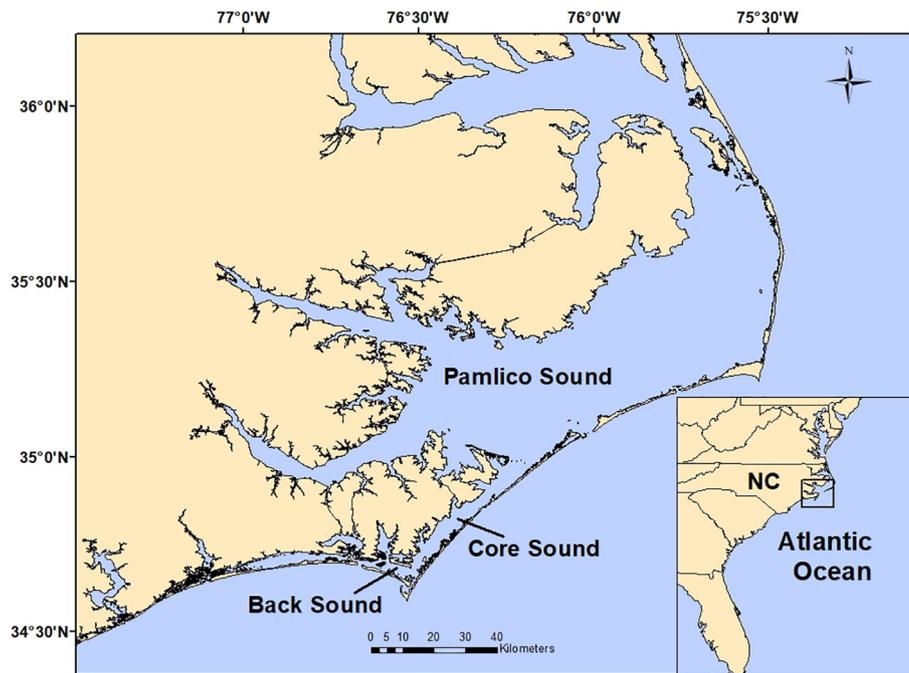


Fig. 1. Study site in waters of North Carolina (NC), USA.

where θ_s is the set of estimated parameters for each sex. The approach can accommodate any suitable functional form of f_s . Here, we consider various polynomials with all or some of the following terms,

$$f_s = \beta_{0,s} + \beta_{1,s}L + \beta_{2,s}T + \beta_{3,s}L^2 + \beta_{4,s}T^2 + \beta_{5,s}L^3 + \beta_{6,s}T^3 \quad (2)$$

where coefficients $\beta = \{\beta_{0,s}, \beta_{1,s}, \dots\}$ are estimated parameters. We applied uniform prior distributions on each member of β , such that $\beta_{0,s} \sim U(-10, 20)$ and $\beta_{1 \dots 6,s} \sim U(-2, 2)$.

Group membership (i.e., sex) of each turtle i was denoted by z_i and modeled as partially observed. If sex of an individual was known, z_i was assigned the value of 1 for males and 2 for females. If unobserved, z_i was treated as a latent variable, assumed to follow a Bernoulli distribution,

$$d_i \sim \text{Bernoulli}(p) \quad (3)$$

where p is the estimated probability of being male, modeled with the hyperprior $p \sim U(0, 1)$. The parameter d_i takes a value of 0 (female) or 1 (male), and thus $z_i = 2 - d_i$.

We considered two error distributions, normal and gamma, to model variability of log testosterone concentration. For the normal distribution,

$$h_i | \beta_s, z_i \sim N(\bar{H}_s, \sigma_s^2) \quad (4)$$

where β_s are the model parameters for each sex Eq. (2), and σ_s is the corresponding and estimated standard deviation. For the gamma distribution,

$$h_i | \beta_s, z_i \sim \text{Gamma}(r_s, \lambda_s) \quad (5)$$

where $r_s = \bar{H}_s^2 / \sigma_s^2$ and $\lambda_s = \bar{H}_s / \sigma_s^2$. For both error distributions, we applied a uniform prior on the standard deviation, $\sigma \sim U(0.1, 10)$.

To implement the model (Supplement B), we used JAGS version 4.2.0 (Plummer, 2003), run in R version 3.4 (R Core Team, 2017) with the R package R2jags (Su and Yajima, 2015). We ran three independent Markov chains, each for 900,000 iterations. Posterior distributions were computed after a burn-in period of 100,000 iterations, and we thinned the resulting chains by keeping every 20th iteration to allow relatively long chains while minimizing computational storage space (Link and Eaton, 2012). Convergence was assessed through visual inspection of trace, density, and autocorrelation plots, and by examining the Brooks-

Gelman-Rubin statistic for values near 1 (Brooks and Gelman, 1998).

2.3. Classification by sex

In each Markov chain Monte Carlo (MCMC) iteration, the model assigns each turtle to a single sex. For an individual with known sex, the assignment is fixed a priori. If unknown, however, the assignment is made probabilistically and may differ across MCMC iterations. We used the mode of the posterior distribution (i.e., the sex assigned most frequently) to classify each turtle, and we used the frequency of the modal assignment to quantify confidence in that classification.

2.4. Model selection

We applied model selection criteria to indicate the optimal error distribution (normal or gamma) and the optimal level of complexity (structure of f_s). Using the training data set, we fit various models defined by Eq. (2), starting from the simplest model (intercept only, $\beta_{0,s}$) and then adding terms while considering whether they improved performance. This procedure was conducted separately for each error distribution Eqs. (4) or (5).

Although numerous model selection criteria have been proposed, none is considered best for all Bayesian applications (Kéry and Schaub, 2012; Hooten and Hobbs, 2015). Thus, we computed three criteria and examined them for consistency: the deviance information criterion (DIC), leave-one-out cross validation (LOO), and the widely applicable information criterion (WAIC, also referred to as Watanabe-Akaike information criterion). DIC (Spiegelhalter et al., 2002) is perhaps the most common criterion and is standard output of JAGS. It has been found to work well in many applications, although it lacks strong theoretical justification (Plummer, 2008; Hooten and Hobbs, 2015). Both LOO (Vehtari and Lampinen, 2002) and WAIC (Watanabe, 2010) have been suggested as reasonable alternatives to DIC (Gelman et al., 2014), and were calculated with the R package loo (Vehtari et al., 2016a; Vehtari et al., 2016b).

2.5. Prediction accuracy

After identifying the optimal model structure, we examined

prediction accuracy for turtles of unknown sex. In this analysis, we used the training data set with known sex ($n = 170$), but successively treated each turtle as if its sex were unknown. Thus, we re-fit the model 170 times, where each iteration contained $n = 169$ turtles of known sex and $n = 1$ turtle of unknown sex. We quantified prediction accuracy by comparing classification of the unknowns to their actual sex.

2.6. Sample size of training data

The training data set provides supervision of the classification algorithm; however, such data are relatively costly. Conducting laparoscopic examination to verify sex requires far more resources (time, funding, equipment, expertise) than does collecting data on other measurable traits, such as carapace length or testosterone concentration. Therefore, we attempted to quantify the improvement in prediction accuracy as a function of sample size in the training data set. The objective was to provide guidance for other, similar studies when evaluating the benefit (classification accuracy) of investment in laparoscopy.

In this analysis, we again utilized the original training data set with $n = 170$ individuals of known sex. From these individuals, we chose at random (without replacement) a subset of sample size $n_t = \{20, 40, 60, \dots, \text{or } 140\}$, and this subset was used as a new training data set. From the remaining $n - n_t$ individuals, we chose at random (without replacement) a second subset of $n_l = 30$ individuals for which sex was treated as unknown. The two subsets were combined to create a data set with sample size $n_t + n_l$. We applied the optimal model structure to the combined data set, re-estimating parameters and computing classification accuracy for the 30 turtles treated as unknown. For each level of n_t , this procedure was repeated 100 times, and we evaluated performance as n_t increased based on the mean and variance of classification accuracy.

2.7. Application to the full data set

The full data set comprised 1401 turtles, including the 170 individuals of known sex. This data set contained only those turtles for which the associated temperature measurement fell inside the range of the training data set, to avoid potentially erroneous extrapolations. In the full data set, the range of straight carapace length was 41.7–75.0 cm, the range of water temperature was 14.0–29.0 °C, and the range of testosterone concentration was 3.4–5560.0 pg/ml (Fig. 2). Turtles of known sex were treated as such by the model, and turtles with unknown sex were classified probabilistically, with these classifications being the primary output of interest. We used a chi-square test to compare our estimated population sex ratio to that from a previous approach in which unknowns were excluded if their testosterone concentration fell within a range where the two sexes overlapped (Braun McNeill et al., 2016).

3. Results

3.1. Fits to the training data

Of the various model configurations fit to the training data set, all three selection criteria indicated that the optimal performance was obtained from the model with an intercept term, a linear relationship with temperature, a quadratic relationship with temperature, and normally distributed error (Table 1). Including carapace length offered no improvement, and therefore that predictor was subsequently removed from consideration. Similarly, including higher order terms did not improve the model; adding a cubic term resulted in poor convergence, because the model was over-parameterized. For the optimal model, posterior median parameter estimates (95% credible intervals) for the male component were $\beta_{0,1} = -1.10$ (-4.48, 2.27), $\beta_{2,1} = 0.64$ (0.30, 0.97), and $\beta_{4,1} = -0.013$ (-0.021, -0.005), and for the female

component were $\beta_{0,2} = -4.93$ (-7.56, -2.21), $\beta_{2,2} = 0.94$ (0.70, 1.20), and $\beta_{4,2} = -0.022$ (-0.028, -0.015). The resulting fits showed considerable curvature in predicted testosterone concentrations as a function of temperature, particularly for females (Fig. 3).

In the “single unknown” analysis, overall prediction accuracy was 94% (160 out of 170 turtles classified correctly; Supplement A). Prediction accuracy for females (95%) was slightly higher than for males (92%). The model predicted with high confidence (> 98% accuracy) that turtles with low testosterone concentrations were female and turtles with high concentrations were male (Fig. 4). The model had lower confidence in predictions at mid-levels of testosterone, where the majority of mis-classifications occurred. Of those turtles classified incorrectly, four were male predicted to be female, and six were female predicted to be male.

3.2. Sample size of training data

As the sample size of the training data set increased, mean classification accuracy of the $n_l = 30$ turtles increased as well, although with diminishing returns (Fig. 5A). At a sample size of $n_t = 20$, mean accuracy was near 88%; at $n_t = 80$, near 92%; and at $n_t = 140$, near 93%. The uncertainty in these values, as indicated by standard errors, was little affected by sample size of the training data set (Fig. 5A). Similarly, the predicted probability (p) of being male and its uncertainty varied little across the range of sample sizes (Fig. 5B).

The primary effect of increased sample size was in the estimated regression coefficients (Fig. 5C–H). Although the median values changed little with increased sample size, the variation in estimates was greatly reduced, indicating greater model stability. However, this effect of sample size was also due to increases in total sample size ($n_t + n_l$), not just that of the training data set alone (n_t).

3.3. Fit to the full data set

The optimal model, when fit to the full data set (Fig. 6), showed similar patterns in the estimated curves as when fit to the training data set. Any differences are due to the influence of turtles with sex treated as a latent variable. For example, the fit to the full data set indicated non-decreasing testosterone concentrations at warmer temperatures, and generally appeared more linear than the fit to the training data set. After observing that result, we re-fit the data with a linear version of the model (i.e., without $\beta_{4,s}$ terms), but found the optimal model with the quadratic term was still favored, as indicated by the information criteria (results not shown). The posterior median parameter estimates (95% credible intervals) for the male component were $\beta_{0,1} = 0.04$ (-2.23, 2.22), $\beta_{2,1} = 0.51$ (0.31, 0.73), and $\beta_{4,1} = -0.009$ (-0.014, -0.005), and for the female component were $\beta_{0,2} = 2.27$ (0.93, 3.62), $\beta_{2,2} = 0.18$ (0.05, 0.30), and $\beta_{4,2} = -0.003$ (-0.006, 0.000). Because $\beta_{0,1} < \beta_{0,2}$, the male curve would eventually decrease below the female curve at lower temperatures, underscoring the caveat not to extrapolate beyond the range of observations. Posterior distributions of estimated parameters are shown in Supplement C.

The posterior median parameter estimate for the probability of being male was $p = 0.17$ (0.14, 0.20). As with the training data set, the model had high confidence in sex classification at low and high concentrations of testosterone (Fig. 7A). Classification confidence ranged from 0.51 to 1.0, with about 93% of all turtles being classified with confidence of at least 0.75 (Fig. 7B). Lower confidence occurred for mid-range concentrations near 400 pg/ml (~6 pg/ml in log space), where the overlap between male and female testosterone was most pronounced (Fig. 2C). Of the 1231 turtles with unknown sex, 211 (17%) were predicted by the model to be male. When considering all 1401 turtles where sex was known ($n = 170$) or predicted ($n = 1231$), the estimated sex ratio was 4.4:1 female (F):male (M). That proportion is significantly higher than 3:1 ($\chi^2 = 12.99$, $df = 1$, $p < .001$), the ratio reported in a previous study where numerous unknowns were excluded

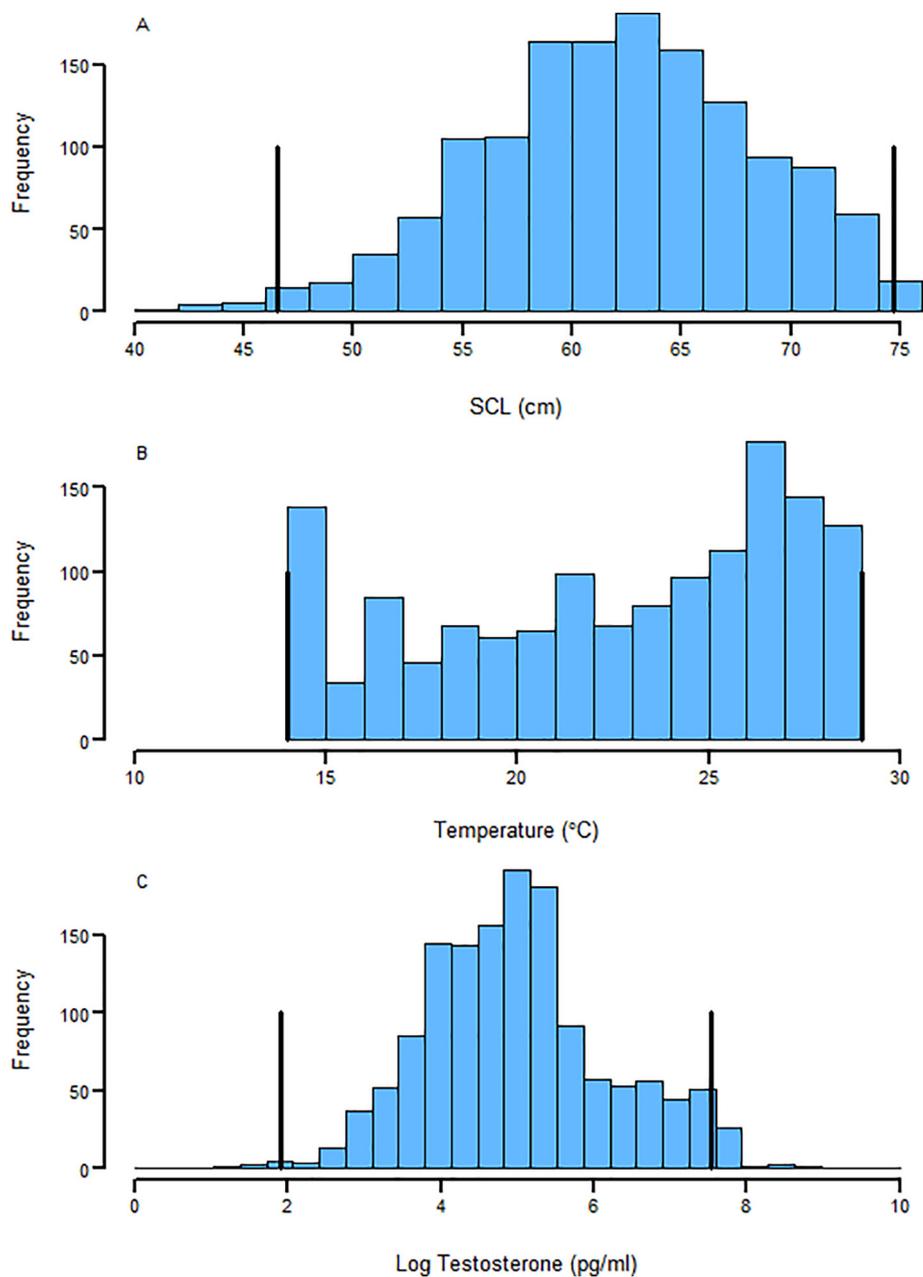


Fig. 2. Distributions of observed values in the juvenile loggerhead dataset, pooled across sexes: (A) straight carapace length (SCL), (B) water temperature, and (C) log of testosterone concentration. Vertical line segments indicate the range of observations in the training data set.

Table 1

Deviance information criterion (DIC), leave-one-out cross validation (LOO), and the widely applicable information criterion (WAIC) for various model configurations Eq. (2) and error structures (Normal or Gamma). Lower values of DIC, LOO, and WAIC indicate better model performance. Bold text denotes best performance of models that converged; italicized text denotes poor convergence of posterior distributions.

Model	Normal			Gamma		
	DIC	LOO	WAIC	DIC	LOO	WAIC
β_0	408.7	409.1	409.0	424.1	425.8	425.7
β_0, β_1	408.9	408.9	408.8	424.7	426.9	426.6
β_0, β_2	383.9	384.5	384.4	399.4	401.9	401.5
$\beta_0, \beta_1, \beta_2$	387.0	387.2	387.1	402.9	405.7	405.1
$\beta_0, \beta_2, \beta_4$	335.5	335.9	335.7	347.3	349.4	348.9
$\beta_0, \beta_2, \beta_4, \beta_6$	335.4	336.2	336.4	347.7	350.6	350.1

from the analysis (Braun McNeill et al., 2016).

4. Discussion

In the current study, we sought to improve the ability to characterize the sex ratio of sea turtle populations using mixture models. Our application utilized the largest published testosterone dataset for juvenile loggerheads, collected during multiple seasons with variable environmental conditions over a continuous 14-year period, 1997–2010. One primary goal was to refine and test analytical methodology for classifying sea turtles by sex based on covariate information. We found that the mixture modeling approach was generally successful, with prediction accuracy of 94%. Although our data set contained information on testosterone concentration, carapace length, and water temperature measurements, the modeling approach is general and could accommodate any observed covariates that might inform classification.

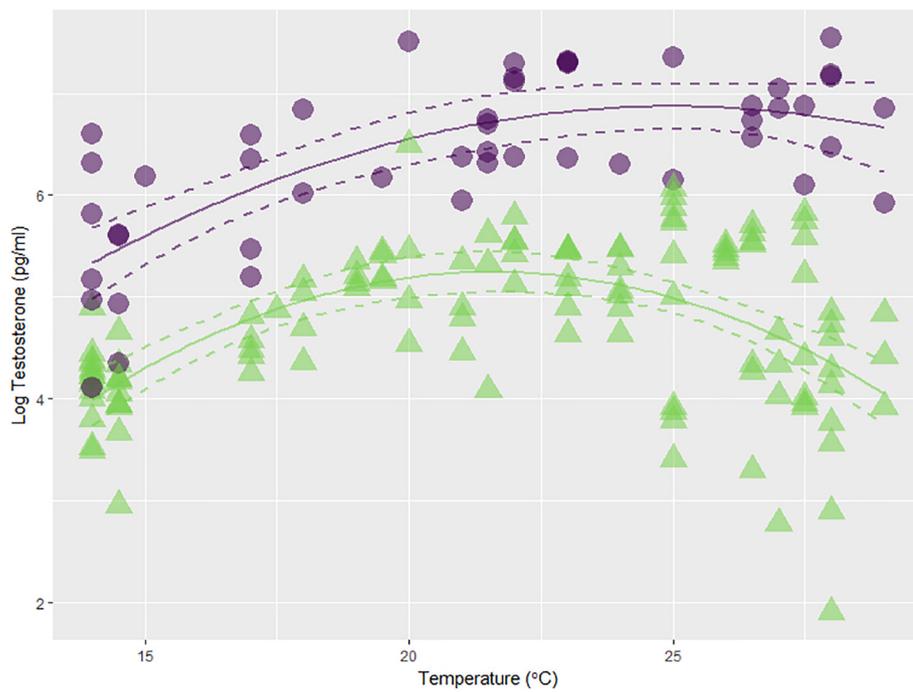


Fig. 3. Fits of the optimal model applied to the training data set in which sex is known. Solid lines represent expected curves based on the medians from the posterior distributions; dashed lines represent 95% credible intervals. Colors and shapes indicate sex, with males (purple, circles) generally having higher testosterone concentrations than females (green, triangles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

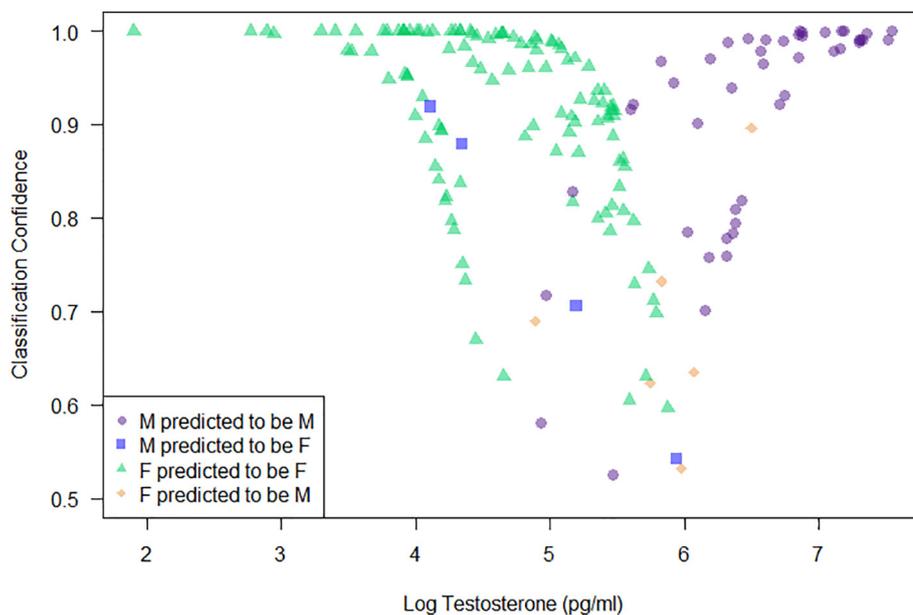


Fig. 4. Prediction accuracy of the optimal model applied to each turtle in the training data set. Purple circles indicate males (M) correctly classified; blue squares indicate males incorrectly classified as female (F); green triangles indicate females correctly classified; and tan diamonds indicate females incorrectly classified as male. Predicted sex is that most often selected in the MCMC iterations, and classification confidence is computed as the proportion of iterations with that prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Another main objective of these analyses was to provide general guidance about sample size for conducting laparoscopy on a subset of turtles from a study population. We found diminishing returns in the sense that classification accuracy saturated as the sample size in the training data set increased (Fig. 5A). Our analysis with this data set suggests that a sample size in the range of 100–140 should be sufficient, somewhat fewer than the sample size of our training data set ($n = 170$). We recommend that similar analyses be conducted on other data sets to test whether this result is robust. In general, we would expect the sufficient sample size to depend on the amount of overlap between male and female testosterone concentrations, with greater overlap requiring more samples to distinguish the sexes, and with greater distinction requiring fewer samples.

We further recommend that efficiency could be gained by targeting turtles for laparoscopy in two ways. First, classification accuracy would likely improve by selecting turtles for laparoscopy that have

intermediate testosterone concentrations in the range where males and females overlap. Recent methods validated for measuring sea turtle testosterone (Allen et al., 2015) have the potential to allow these data to be available quickly (within a day), facilitating optimal selection of individuals for laparoscopy. Second, incorporating data from turtles sampled over a broad range of temperatures would help inform estimation of the model's regression coefficients. Including temperatures at the extremes (low and high) would be particularly useful, to avoid the temptation to extrapolate beyond the range of observations.

The saturation in classification accuracy occurred because the covariates themselves are informative for separating the components (sexes) of the mixture model. The primary effect of the training data set was to order the regression polynomials (f_1 and f_2) during the MCMC procedure. This avoided a known ambiguity when fitting mixture models: parameters are not identifiable if the mixture distribution remains unchanged when the components' labels are permuted

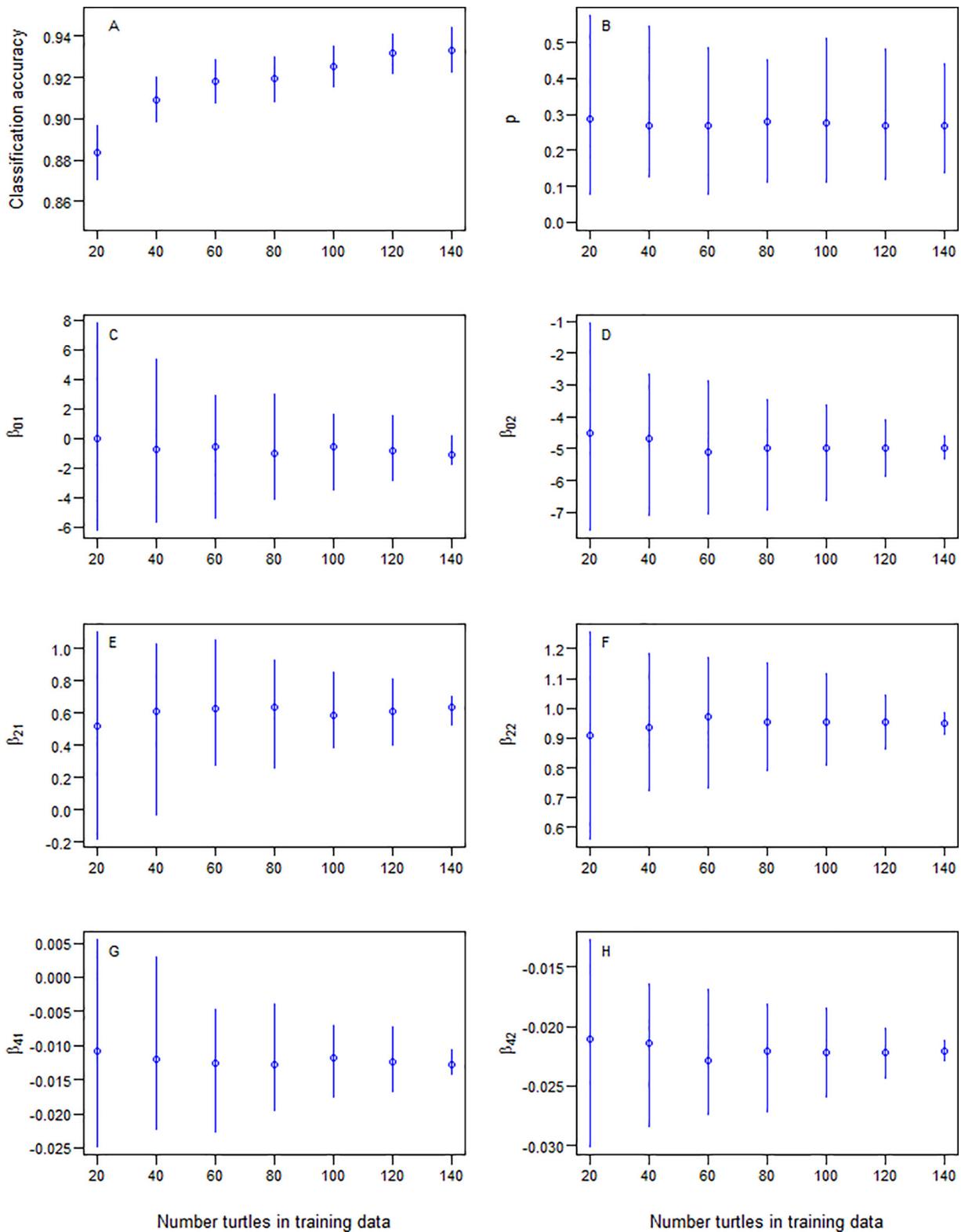


Fig. 5. Effects of sample size of the training data set on (A) mean classification accuracy (\pm two standard errors) of 30 turtles with sex treated as unknown, (B) median estimated probability (p) of being male with the line representing 2.5th to 97.5th percentiles, and (C–H) median estimated regression coefficients (see text) with the lines representing 2.5th to 97.5th percentiles. Means, medians, and other percentiles were computed from 100 randomized replicates.

(McLachlan and Peel, 2000; Gelman et al., 2013). In our model, males were labeled as component one ($s = 1$) and females as component two ($s = 2$), however the MCMC procedure would be blind to this definition without the training data set, such that two MCMC chains could achieve identical regression coefficients, but with the labels reversed. In our full

data set of 1401 turtles, those of known sex comprised only about 12% of the total sample size, but this was sufficient during estimation to order the components of the mixture distribution.

Previous studies have considered testosterone concentrations when examining the sex ratio of this loggerhead population (Braun McNeill

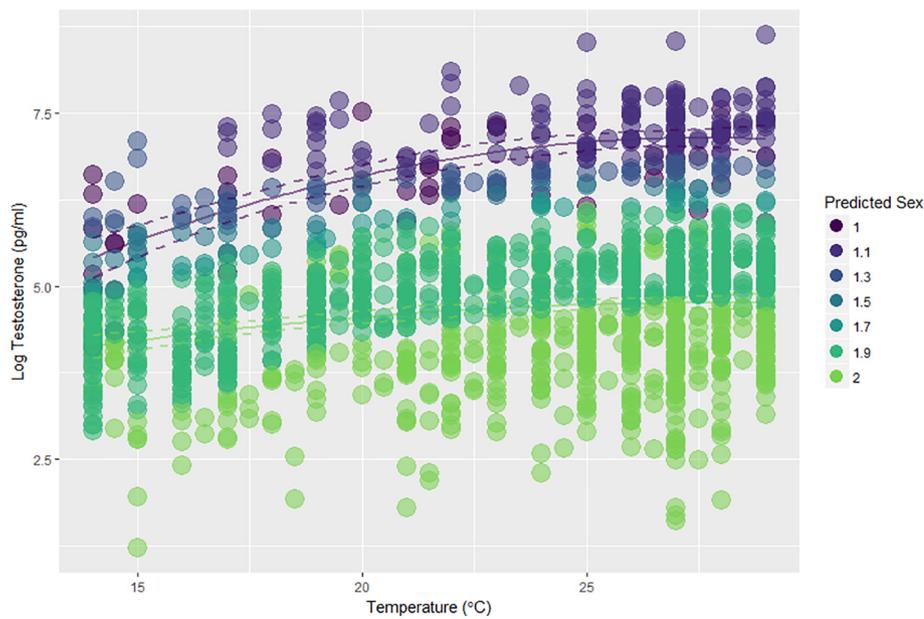


Fig. 6. Fits of the optimal model applied to the full data set. Solid lines represent expected curves based on the medians from the posterior distributions; dashed lines represent 95% credible intervals. Colors indicate the gradient of prediction that an individual is male (Predicted Sex = 1) or female (Predicted Sex = 2), calculated as the posterior mean value of Bernoulli variable z_i .

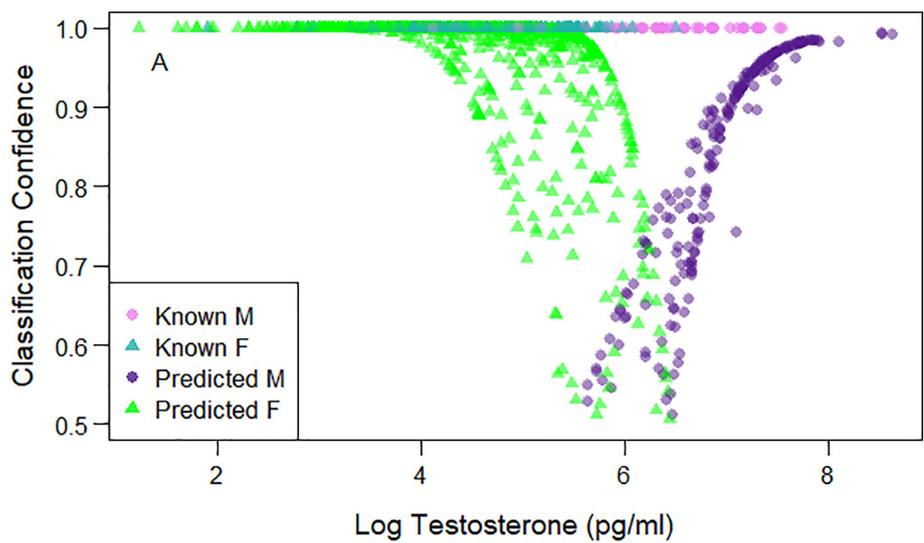
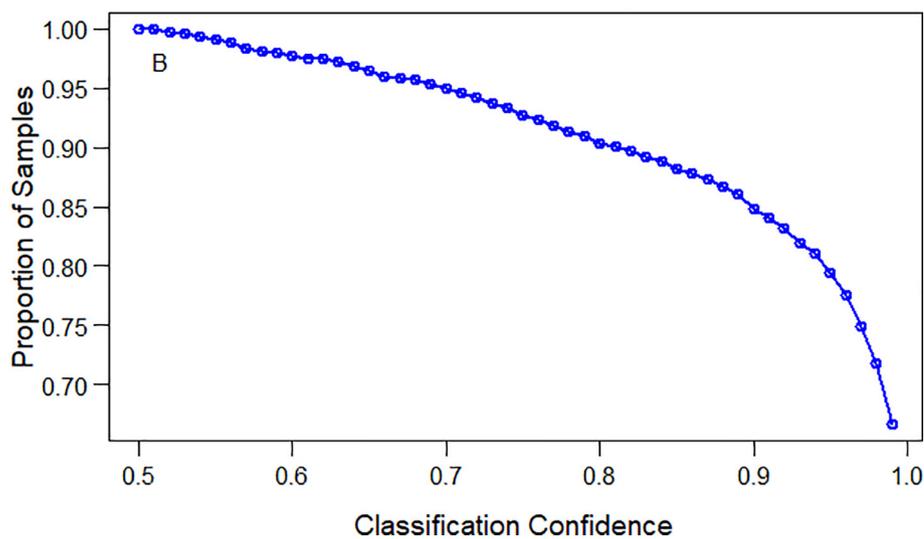


Fig. 7. Classification confidence of the optimal model applied to each turtle in the full data set (A). Violet circles indicate males (M) known to be male (from the training data set); light blue triangles indicate females (F) known to be female (from the training data set); purple circles indicate turtles of unknown sex classified as male; and green triangles indicate turtles of unknown sex classified as female. Predicted sex is that most often selected in the MCMC iterations, and classification confidence is computed as the proportion of iterations with that prediction. Panel (B) shows the proportion of samples (turtles) in the data set meeting or exceeding various levels of classification confidence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



et al., 2007; Braun McNeill et al., 2016). However, those studies did not explicitly account for effects of covariates, such as temperature, but rather excluded turtles sampled at < 20 °C, based on indications that testosterone concentrations were depressed at lower water temperatures. In addition, those studies excluded data where testosterone concentrations of males and females overlapped. The mixture modeling approach here improves on those previous studies by including the effect of temperature and by predicting sex in the region of overlap, increasing the available sample size for the study population from 901 (Braun McNeill et al., 2016) to 1401. Furthermore, the approach quantifies the uncertainty of predictions in the form of classification confidence (Fig. 7).

Previous analyses indicated that the sex ratio for this juvenile loggerhead population was 3.0 F:1.0 M, i.e. 75% female (Braun McNeill et al., 2016). However, applying the approach described herein to the full data set yielded a sex ratio of 4.4 F:1.0 M, with 81% of turtles estimated to be female. These sex ratios calculated using the two methods are significantly different. This discrepancy highlights the potential influence of turtles categorized as “unknown” due to intermediate testosterone concentrations and/or the influence of covariates such as water temperature on the ability to characterize population sex ratios accurately.

Given the high conservation concerns for sea turtles worldwide, understanding their demography is critical. These concerns are amplified by the potential for climatic warming to increase egg incubation temperatures and thereby affect sex ratios. Thus, it is vital to apply research methods that provide the best possible information, while minimizing negative impacts of sampling. The results of this study provide a means to refine sampling and improve analytical methods for estimating sex ratios of sea turtle populations, enhancing our ability to support management and conservation efforts.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jembe.2018.03.006>.

Acknowledgements

We are grateful for the constructive guidance of D. Owens; for the radioimmunoassays conducted by G. Blanvillain, L. Evich, M. Lee, R. Patterson, and J. Schwenter; for laparoscopies conducted by B. Chittick, G. Lewbart, A. Segars, A. Stamper, S. Weber, and S. Willens; for assistance with fieldwork and laparoscopies from many volunteers; and for the cooperation of numerous commercial fishermen. We thank J. Schwenter for providing us with quality assurance information on the radioimmunoassay procedure, and we thank T. Eguchi, M. Godfrey, E. Williams, and anonymous reviewers for critical comments on the manuscript. Opinions expressed are those of the authors and do not necessarily represent views of any government agency.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Allen, C.D., Robbins, M.N., Eguchi, T., Owens, D.W., Meylan, A.B., Meylan, P.A., Kellar, N.M., Schwenter, J.A., Nollens, H.H., LeRoux, R.A., Dutton, P.H., Seminoff, J.A., 2015. First assessment of the sex ratio for an east Pacific green sea turtle foraging aggregation: validation and application of a testosterone ELISA. *PLoS One* 10 (10), e0138861.

Arendt, M.D., Schwenter, J.A., Segars, A.L., Byrd, J.L., Maier, P.P., Whitaker, J.D., Owens, D.W., Blanvillain, G., Quattro, J.M., Roberts, M.A., 2012. Catch rates and demographics of loggerhead sea turtles (*Caretta caretta*) captured from the Charleston, South Carolina, shipping channel during the period of mandatory use of turtle excluder devices (TEDs). *Fish. Bull.* 110, 98–109.

Avens, L., Goshe, L.R., Coggins, L., Snover, M.L., Pajuelo, M., Bjorndal, K.A., Bolten, A.B., 2015. Age and size at maturation and adult stage duration for loggerhead sea turtles in the western North Atlantic. *Mar. Biol.* 162, 1749–1767.

Benoit, H.P., Hurlbut, T., Chasse, J., Jonsen, I.D., 2012. Estimating fishery-scale rates of

discard mortality using conditional reasoning. *Fish. Res.* 125–126, 318–330.

Blanvillain, G., Wood, L.D., Meylan, A.B., Meylan, P.A., 2008. Sex ratio prediction of juvenile hawksbill sea turtles (*Eretmochelys imbricata*) from South Florida, USA. *Herpetol. Conserv. Biol.* 3, 21–27.

Bolker, B.M., Toshinori, O., Bjorndal, K.A., Bolten, A.B., 2007. Incorporating multiple mixed stocks in mixed stock analysis: ‘many-to-many’ analyses. *Mol. Ecol.* 16, 685–695.

Bolten, A.B., Bjorndal, K.A., Grumbles, J.S., Owens, D.W., 1992. Sex ratio and sex-specific growth rates of immature green turtles, *Chelonia mydas*, in the southern Bahamas. *Copeia* 1992, 1098–1103.

Bolten, A.B., Bjorndal, K.A., Martins, H.R., Dellinger, T., Biscoito, M.J., Encalada, S.E., Bowen, B.W., 1998. Transatlantic developmental migrations of loggerhead sea turtles demonstrated by mtDNA sequence analysis. *Ecol. Appl.* 8, 1–7.

Braun McNeill, J., Epperly, S.P., Owens, D.W., Avens, L., Williams, E., Harms, C.A., 2007. Seasonal reliability of testosterone radioimmunoassay (RIA) for predicting sex ratios of juvenile loggerhead (*Caretta caretta*) turtles. *Herpetologica* 63, 275–284.

Braun McNeill, J., Avens, L., Goodman Hall, A., Goshe, L.R., Harms, C.A., Owens, D.W., 2016. Female-bias in a long-term study of a species with temperature-dependent sex determination: monitoring sex ratios for climate change research. *PLoS One* 11 (8), e0160911. <http://dx.doi.org/10.1371/journal.pone.0160911>.

Brooks, S., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7, 434–455.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Davenport, J., 1997. Temperature and the life-history strategy of sea turtles. *J. Therm. Biol.* 22, 479–488.

Geis, A., Wibbels, T., Phillips, B., Hillis-Starr, Z., Meylan, A., Meylan, P., Diez, C., Van Dam, R., 2003. Predicted sex ratio of juvenile hawksbill sea turtles inhabiting Buck Island Reef National Monument, U.S. Virgin Islands. *J. Herpetol.* 37, 400–404.

Geis, A.A., Barichivich, W.J., Wibbels, T., Coyne, M., Landry Jr., A.M., Owens, D., 2005. Predicted sex ratio of juvenile Kemp's ridley sea turtles captured near Steinhatchee, Florida. *Copeia* 2005, 393–398.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. *Bayesian Data Analysis*, 3rd Ed. CRC Press, Boca Raton, FL.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* 24, 997–1016.

Guthrie, J.F., Kroger, R.L., Gordy, H.R., Lewis, C.W., 1973. The long-haul fishery of North Carolina. *Mar. Fish. Rev.* 35, 27–33.

Hamann, M., Godfrey, M.H., Seminoff, J.A., et al., 2010. Global research priorities for sea turtles: informing management and conservation in the 21st century. *Endanger. Species Res.* 11, 245–269.

Hamel, S., Yoccoz, N.G., Gaillard, J., 2017. Assessing variation in life-history tactics within a population using mixture regression models: a practical guide for evolutionary ecologists. *Biol. Rev.* 92, 754–775.

Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Ed. Springer-Verlag, New York.

Hawkes, L.A., McGowan, A., Godley, B.J., Gore, S., Lange, A., Tyler, C.R., Wheatley, D., White, J., Witt, M.J., Broderick, A.C., 2013. Estimating sex ratios in Caribbean hawksbill turtles: testosterone levels and climate effects. *Aquat. Biol.* 18, 9–19.

Hays, G.C., Mazaris, A.D., Schofield, G., 2014. Different male vs. female breeding periodicity helps mitigate offspring sex ratio skews in sea turtles. *Front. Mar. Sci.* <http://dx.doi.org/10.3389/fmars.2014.00043>.

Heppell, S.S., Snover, M.L., Crowder, L.B., 2003a. Sea turtle population ecology. In: Lutz, P.L., Musick, J.A., Wyneken, J. (Eds.), *The Biology of Sea Turtles*. Vol. II. CRC Press, Boca Raton, Florida, USA, pp. 275–306.

Heppell, S.S., Crouse, D.T., Crowder, L.B., Epperly, S.P., Frazer, N.B., 2003b. Population models for Atlantic loggerheads. In: Bolten, A.B., Witherington, B.E. (Eds.), *Synopsis of the Biology and Conservation of the Loggerhead Sea Turtles*. Smithsonian Institution Press, Washington DC, USA, pp. 255–273.

Higgins, E., Pearson, J.C., 1928. Examination of the summer fisheries of Pamlico and Core Sounds, N.C., with special reference to the destruction of the undersized fish and the protection of the gray trout *Cynoscion regalis* (Block and Schneider), Appendix II. In: Report of the United States Commissioner of Fisheries for the Fiscal Year 1927 with Appendices. US Department of Commerce, Bureau of Fisheries, Government Printing Office, Washington D.C.

Hooten, M.B., Hobbs, N.T., 2015. A guide to Bayesian model selection for ecologists. *Ecol. Monogr.* 85, 3–28.

Kéry, M., Schaub, M., 2012. *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*. Academic Press, Waltham, MA.

Laloë, J., Cozens, J., Renom, B., Taxonera, A., Hays, G.C., 2014. Effects of rising temperature on the viability of an important sea turtle rookery. *Nat. Clim. Change.* <http://dx.doi.org/10.1038/nclimate2236>.

León, Y.M., Diez, C.E., 1999. Population structure of hawksbill turtles on a foraging ground in the Dominican Republic. *Chelonian Conserv. Biol.* 3, 230–236.

Link, W.A., Eaton, M.J., 2012. On thinning of chains in MCMC. *Methods Ecol. Evol.* 3, 112–115.

Maclean, R.A., Harms, C.A., Braun-McNeill, J., 2008. Propofol anesthesia in loggerhead (*Caretta caretta*) sea turtles. *J. Wildl. Dis.* 44, 143–150.

McLachlan, G., Peel, D., 2000. *Finite Mixture Models*. John Wiley & Sons, Inc., New York, NY.

Millar, R.B., 1987. Maximum likelihood estimation of mixed stock fishery composition. *Can. J. Fish. Aquat. Sci.* 44, 583–590.

Owens, D.W., 1997. Hormones in the life history of sea turtles. In: Lutz, P.L., Musick, J.A. (Eds.), *The Biology of Sea Turtles*. CRC Press, Boca Raton, Florida.

Owens, D.W., Ruiz, G.J., 1980. New methods of obtaining blood and cerebrospinal fluid. *Herpetologica* 36, 17–20.

- Owens, D.W., Hendrickson, J.R., Lance, V., Callard, I.P., 1978. A technique for determining sex of immature *Chelonia mydas* using a radioimmunoassay. *Herpetologica* 34, 270–273.
- Patterson, T.A., Basson, M., Bravington, M.V., Gunn, J.S., 2009. Classifying movement behavior in relation to environmental conditions using hidden Markov models. *J. Anim. Ecol.* 78, 1113–1123.
- Pella, J., Masuda, M., 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fish. Bull.* 99, 151–167.
- Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Hornik, K., Leisch, F., Zeileis, A. (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Austria, Vienna.
- Plummer, M., 2008. Penalized loss functions for Bayesian model comparison. *Biostatistics* 9, 523–539.
- Rees, A.F., Alfaro-Shigueto, J., Barata, P.C.R., et al., 2016. Are we working towards global research priorities for management and conservation of sea turtles? *Endanger. Species Res.* 31, 337–382.
- Sanchez, C.L., 2013. Sex Ratios of Juvenile Green Turtles (*Chelonia mydas*) in three Developmental Habitats along the East Coast of Florida. Thesis. University of Central Florida.
- Shertzer, K.W., Fieberg, J., Potts, J.C., Burton, M.L., 2017. Identifying growth morphs from mixtures of size-at-age data. *Fish. Res.* 185, 83–89.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. B.* 64, 583–639.
- Su, Y.S., Yajima, M., 2015. R2jags: Using R to Run 'JAGS'. R package version 0.5–7. <https://cran.r-project.org/package=R2jags>.
- Sweeney, K.L., Shertzer, K.W., Fritz, L.W., Read, A.J., 2015. A novel approach to compare pinniped populations across a broad geographic range. *Can. J. Fish. Aquat. Sci.* 72, 175–185.
- Vehtari, A., Lampinen, J., 2002. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput.* 14, 2439–2468.
- Vehtari, A., Gelman, A., Gabry, J., 2016a. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. arXiv preprint arXiv:1507.04544.
- Vehtari, A., Gelman, A., Gabry, J., 2016b. Loo: Efficient Leave-one-out cross-Validation and WAIC for Bayesian Models. R package version 0.1.6. <https://github.com/jgabry/loo>.
- Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* 11, 3571–3594.
- Wibbels, T., 1988. Gonadal Steroid Endocrinology of Sea Turtle Reproduction. Texas A&M University (Dissertation).
- Wibbels, T.R., 1999. Diagnosing the sex of sea turtles in foraging habitats. In: Eckert, K.L., Bjørndal, K.A., Abreu-Grobois, F.A., Donnelly, M. (Eds.), *Research and Management Techniques for the Conservation of Sea Turtles*.
- Wibbels, T.R., 2003. Critical approaches to sex determination in sea turtles. In: Lutz, P.L., Musick, J.A., Wyneken, J. (Eds.), *The Biology of Sea Turtles*. Vol. II. CRC Press, Boca Raton, Florida, USA, pp. 103–134.
- Wibbels, T., Owens, E.W., Morris, Y., Amoss Jr., M.S., 1987. Seasonal changes in the serum testosterone titers of loggerhead sea turtles captured along the Atlantic coast of the United States. In: Witzell, W. (Ed.), *Ecology of East Florida Sea Turtles*, Book NOAA Technical Report NMFS-53.
- Wibbels, T.R., Martin, R.E., Owens, D.W., Amoss Jr., M.S. 1991. Female-biased sex ratio of immature loggerhead sea turtles inhabiting the Atlantic coastal waters of Florida. *Can. J. Zool.* 69, 2973–2977.
- Wibbels, T., Owens, D.W., Limpus, C.J., 2000. Sexing juvenile sea turtles: is there an accurate and practical method? *Chelonian Conserv. Biol.* 3, 756–761.
- Witzell, W.N., Geis, A.A., Schmid, J.R., Wibbels, T., 2005. Sex ratio of immature Kemp's ridley turtles (*Lepidochelys kempi*) from Gullivan Bay, Ten Thousand Islands, south-west Florida. *J. Mar. Biol. Ass. U.K.* 85, 205–208.
- Wood, J.R., Wood, F.E., Critchley, K.H., Wildt, D.E., Bush, M., 1983. Laparoscopy of the green sea turtle, *Chelonia mydas*. *Br. J. Herpetol.* 6, 232–237.