Biological Data Collection Methods and Quantitative Analyses for the Assessment of the Hawaiian stock of the Green Sea Turtle (Chelonia mydas): CIE Review

Michael C.S. Kingsley

May 2006

Report prepared for the University of Miami Independent System for Peer Review.

Contents

Executive Summary

Kingsley, M.C.S. 2006. Biological Data Collection Methods and Quantitative Analyses for the Assessment of the Hawaiian stock of the Green Sea Turtle (*Chelonia mydas*): CIE Review. Rep. prep. for the Univ. of Miami Independent System for Peer Review. Unpublished. iv+27 pp.

The Statement of Work for this review asks for attention to: field methods for nesting beach research; methods for at-sea research, particularly in respect of the study design and its ability to estimate stock-dynamic parameters at the population level; and methods used for data analysis, including analysis of nesting-beach surveys, population trend analysis, growth analyses, and the modelling of stock dynamics, using Bayesian methods, for stock assessment.

A review meeting held in Honolulu 9–10 May 2006 was unfortunately not well aligned with the objectives of the review. There was little emphasis on presenting quantitative material on methods used for analysing data or on the sampling design of the at-sea research.

The field methods used to estimate annual numbers of nesters at East Island appear unexceptionable. Sampling of the island might at some point become appropriate if numbers continue to increase. There would be an advantage in periodically repeating surveys that extend over the entire nesting season to re-calibrate the partial-season surveys, and in confirming the validity of using the East Island numbers as an index for the entire Hawaiian stock by periodically surveying other breeding beaches.

In most of the documents reviewed, the methods used to analyse the data are described only in very general terms, models are not specified by equations, criteria for model fit are poorly, or not, specified. There is a frequent use of non-parametric models that, generally speaking, fit the data with smooth curves that are not quantitatively defined.

Quantitative results, where they exist are poorly presented, often only in graphical form; presentations of results are often incomprehensible, and measures of model fit are in most cases not given. Bayesian methods are referred to, but with no definition of prior distributions or demonstration that the priors are updated by the data. The methods might be appropriate, but on the information available it is difficult to evaluate how well they have been applied or the usefulness of the results.

While results in general are at least plausible and may well be qualitatively correct, in nearly all cases where checks or re-analyses were possible there were errors in methods, analyses, or interpretation of results, in the direction of the denial of uncertainty. The uncertainty of East I. nest numbers is calculated by a method that (probably) underestimates it; precision of the estimate of the rate of increase of those numbers is exaggerated by a factor of about 4; uncertainty in the rate of increase due to nesting habitat changes is not discussed; the inability of the assessment model to provide estimates of stock-dynamic parameters is simply ignored.

Making reasonable assumptions about sources of uncertainty, the methods used for analyzing the East Island nester survey data and for estimating the uncertainty of partial-survey estimates of nester numbers appear likely to overstate the precision of the estimates—perhaps by a large factor—and to bias the estimates upwards by some small amount. The published analysis of the 30-year series of nester estimates at East Island appears faulty: the amplitude of the extracted cyclical component is too small by half, the rest having been left in the reported residuals; and the estimate of the rate of increase is about 4 times less precise than it is made out to be. Publishing the trend in East Island nester numbers as though it applied to the entire population without at all discussing the possible effects of the changes in nesting habitat (including the disappearance of entire islands?) that have occurred in the archipelago appears scientifically unsound.

Information on the sampling design of the at-sea programme was not readily forthcoming. It appears that the objective of the programme has always been chiefly to collect longitudinal data on individuals. To that end, I gathered, sampling is concentrated in certain areas in the coastal pastures that are repeatedly visited. The programme successfully reaches its objective and has produced an abundance of longitudinal data. But such a concentration of the sampling might well unfit the programme for producing estimates of population-level stock-dynamic parameters. This is not a simple issue, but might be resolved by a workshop focused on this one subject and considering, whether the at-sea sampling programme would have to be changed to estimate stock-dynamic parameters (and if so how), and how successfully it could do so even if it were to be altered in that direction.

The attempt to fit an assessment model by Bayesian methods presents serious problems. Equations defining the model are, as they are reported, wrong. The data is inadequate to allow such a model to be fitted, and the results of the Bayesian analysis confirm this: the prior distributions postulated are not updated in fitting the model to the data, but left unchanged. However, the inability, and failure, of the data to define the carrying capacity, the shape of the stock-recruitment curve, or the present status of the stock are glossed over. Regardless of their shortcomings, the unsubstantiated—unsubstantiable—estimates of stock-dynamic parameters are used as a basis for forecasts of population trajectory under exploitation. These forecasts have little validity.

With stock assessment and management in mind, possible extensions of the research programme might include: repeating full-season surveys at East Island; surveying other breeding beaches; surveys of turtle numbers on the coastal pastures; field studies of the productivity and condition of the vegetation on the pastures.

1. Statement of Work.

The objective of the review as given in the Statement of Work is quite clear: 'a thorough examination of the appropriateness of data collection methods and analytical techniques used for Hawaiian green turtle recovery status and population monitoring', accompanied by a more specific requirement to analyse published documents on the population, focussing on the quantitative methods used for nesting beach research, on the experimental design and data collection methods for the marine habitat research, and on the analytical techniques, with emphasis on Bayesian state-space modelling.

This requirement does not include consideration of the content of the research programme. However, it seems appropriate to include a few comments on that, for two reasons: one is that the content of the research and monitoring programme is relevant to the intention of the review, which is presumed to be ensuring an adequate monitoring programme; the other is that the review meeting in Honolulu was far more preoccupied with the content of the programme than with the adequacy of the methods used for analysing the data that it gathers.

2. Research and Monitoring Programme

2.1 The main components of the research and monitoring programme are:

- annual monitoring of the number of turtles nesting on East Island, French Frigate Shoals. Surveys are of two types, either 'saturation surveys' in which the nesting beaches are observed for the entire nesting season, or 'partial surveys' in which observations are made only for a shorter period within the nesting season. Females arriving to nest are also tagged with (now) subcutaneous transponder tags.
- 'at-sea' monitoring, which (apparently) consists of capturing, measuring and tagging turtles on the near-shore pastures round the coasts of the main Hawaiian islands;
- the collection and examination of dead, moribund, sick or injured turtles found on land; examination may include up to a full veterinary necropsy, or may be limited by

the condition of the carcass. The stranding programme includes a voluntary component of treatment (including surgery), rehabilitation and release of some sick or injured turtles.

- investigation of the aetiology of a prevalent disease, a fibropapillomatosis of presumably (and presumed) viral origin; the disease is chronic, in nearly all cases progressive, debilitating and also incapacitating in other ways, and probably in most cases eventually fatal;

- DNA (apparently hitherto mostly mtDNA) analysis of samples collected in the course of other studies to elucidate population structure of the species globally, but more especially within the Pacific basin;

- some satellite tagging/tracking of turtles handled in the course of other studies to investigate possible connections between apparently different stocks.

2.2 The history of the programme seems to be that it has built up on an opportunistic basis, collecting data of kinds that were readily collected, and at times and in places where turtles were easily accessible. In other words, the programme has not been designed to answer specific questions, more to collect such data as was there to be collected. This has had two consequences. The first—resoundingly positive—is that the data collection effort has been readily sustained, with the result that long series of data collected over time by the same people using the same methods have become available. The value of long series of consistent data is universally acknowledged. However, the second consequence is that the data collected have not necessarily been well suited to answering specific questions that have been raised; data collection of this kind can suffer from flaws in design that do not get rectified, because the lack of an investigative impetus means that analysis is deferred.

Also, the lack of a driving question can lead to delays in publication, and to publications of a kind that are more directed to analysing and publishing a particular data set than to answering a priority question.

2.3 Future developments of the research and monitoring programme.

Specifically in respect to the monitoring of population status with respect to benchmarks, the following might be considered as possible additions to a monitoring programme:

a. surveys of numbers in the pasture areas. I don't know how such surveys might be designed or executed, whether from the air, from small boats, or by diver, by strip or line transect or by point observations. However, this is a late-maturing species so the mature females counted on the nesting beaches are only a small fraction of the population, and problems with fertility or juvenile survival would not quickly be detected by nesting-beach surveys. For estimating numbers in large populations, the mathematics goes against mark-recapture methods and swings in favour of surveys.

b. extension and perhaps re-design of the at-sea sampling, with a view to assessing whether it can come to answer questions about population-level parameters;

c. study of the condition of the range in the pasture areas, including perhaps installation of enclosures to monitor the effect of grazing by turtles on the standing stocks of seaweeds and sea-grasses. Suggestions that the population may be approaching the carrying capacity of its habitat should indicate that effects on the range could be observed. We heard that some transect surveys of vegetation have been, or are to be, started in a restricted area and on a pilot basis, but without much detail.

d. the saturation surveys on the nesting beaches at East Island probably ought to be repeated at some intervals, perhaps in two consecutive years every 10 or 12. The point was made in the meeting that much is being taken for granted in assuming that the temporal distribution of breeding activity is not changing or has not changed.

e. other nesting islands should be surveyed, at least at intervals. The trend at East I. is assumed to be the trend for the population, and this might be true—but then it also might not, especially when whole islands have disappeared within the period. And obviously, reporting a 3.6% e.c.v.[1] for the estimate of the trend at East Island—even if it were correct—has little point in the context of the population as a whole, especially if the habitat elsewhere is changing.

---

[1] e.c.v.—error coefficient of variation.

f. post-hatch mortality in waters near the nesting beaches was mentioned as a possible subject for study, but I think that mortality elsewhere or at other periods of juvenile life is likely to be so large as to make this fairly pointless. Developing study methods on the coastal pastures in the main islands to be able quantitatively to monitor recruitment of small subadults is likely to be more profitable than attempting to measure post-hatch survival near the breeding beaches.

3. Review Process.

The review was composed of a set of documents, including both published and draft documents, and a 2-day review meeting. At the review meeting, the following presentations were given:

Dr. Bud Antonelis – Welcome, introductions and overview of the purpose, scope and goals of the review – Statement of Work;
George Balazs – Overview of the Marine Turtle Research Program with emphasis on the Hawaiian green turtle metapopulation;
Dr. Peter Dutton – Genetics research of the Hawaiian green turtle;
Dr. Thane Wibbels – Sex determination and pivotal temperature research of the Hawaiian green turtle;
Dr. Jerry Wetherall – Development of quantitative techniques for monitoring and sampling numbers and trends of nesting green turtles at the Hawaiian green turtle rookery of French Frigate Shoals;
Dr. Milani Chaloupka – Analyses of research data for the Hawaiian green turtle metapopulation;
Marc Rice – NOAA/HPA collaborative field studies on the Hawaiian green turtle
George Balazs – Hawaiian sea turtle stranding research network;
Dr. Thierry Work – Fibropapilloma research and health assessments of the Hawaiian green turtle;
Dr. Robert Morris – Diagnosis, treatment and rehabilitation of stranded green turtles in Hawaii;

Some of these presentations were more relevant to the purpose of the review—as described in the Statement of Work—than others. In general, it seemed that the arrangement of the meeting was not well aligned with the scope and objective of the review. Specifically, the only presentation that gave a reasonable treatment of methods for data collection or analysis was Jerry Wetherall's. Milani Chaloupka's presentation gave a very cursory review of results, only, from the analysis of stranding and other data, and most of the other presentations were strongly results-oriented. Overall, there was very little presentation of methods of the analysis of data. Also lacking was any description of the basic methods of the 'at-sea' research effort, such as the sampling design, the geographical distribution of the effort or the samples, the objectives of the at-sea work, &c. It was almost as though the people who asked for the review and the people who organised the meeting were different people with different interests.

4. Response to the requests in the Statement of Work.

4.1. Analytical methods—general.

Analytical methods are tools in a tool-box. In getting satisfactory results, the tool itself is only part of the story—the skill of the user has also a part to play. Sophisticated tools are nice to have, if one knows how to use them, but it's a lot easier to cut one's foot off with a chain-saw than an ordinary hand-saw.

In the documents reviewed, with only one or two exceptions, the descriptions of the methods used for the analysis of data are cursory in the extreme. No models are fully described, certainly not by equations, and the techniques used are named, usually, only in the most general terms, while there is extensive text on irrelevancies such as the computer packages used. This has not facilitated the execution of the SOW. The methods, generally, are appropriate to the objective, which seems to be to get a not-very-quantitative article into some not-very-quantitative journal or other. But the devil's in the details; and the details aren't there. There is heavy use of multivariate non-parametric model fitting, a useful technique for putting smooth curves simultaneously in several dimensions through data. Results are presented generally in graphic forms capable of qualitative interpretation—tabulation of quantitative results is generally absent. This is all probably adequate to journal publication, but for serious decision-making a more quantitative approach would be necessary, including more detailed statements of the quality of model fits.

An exception to the general lack of quantitative treatment is the document on the state-space assessment model fitted by Bayesian methods. In this document there is a partial description of the models used, and a moderately full tabulation of quantitative results.

4.2 An evaluation of the appropriateness of the quantitative methods being used for nesting beach research.

I'm no expert on counting turtles and can't comment on the appropriateness of the field methods, but provided that the counting of emergences and the identification of individual turtles can be accepted as practicable, these methods should give valid usable data. Improving the efficiency of the East Island survey by restricting the survey period is acceptable—far more uncertainty about the state of the stock is probably being incurred by continuing to rely on one set of saturation surveys and by ignoring what is happening elsewhere in the archipelago. If the population expands so much that it becomes impracticable to execute the programme on the whole of East Island, it could be appropriate to sample the island; it is after all itself only an arbitrary sample of the nesting habitat of the population. If this were to be foreseen, it would be appropriate to prepare for it by defining sampling strata on the island.

The analysis methods for estimating the population could be improved. My analysis would be as follows:

For each saturation-survey year j we have a total number of nesters N_j, all individually identified and recognised on each emergence. We also have an 'activity matrix'—a complete record of the nights on which each turtle came ashore, throughout the season—which enables us to know which turtles emerged within any given period.

For each partial-survey year i we have a survey period T_i defined by start and end dates; we have a number of emergences m_i and a number of encountered nesters s_i.

For each combination of a saturation survey j and a partial survey i we can count the number of emergences m'_{ij} that occurred on the saturation survey within the period T_i and estimate a number of nesters by assuming the nesting activity has the same relative intensity over time in all years.

$$\hat{n}_{ij} = \frac{m_i}{m'_{ij}} N_j$$

If there are J saturation surveys giving independent estimates of the ratio N_j/m'_{ij}, this gives J independent estimates of n'_{ij}, from which we can get a mean

$$\hat{n}_i = \sum_j \hat{n}_{ij} \Big/ J$$

And the uncertainty of n_i can be obtained in the usual way by comparing the J different values. The uncertainty attached to estimates of differences between the n_i is probably affected by error correlations between them due to their being based on the same set of saturation surveys—trends in numbers are likely to be estimated more precisely than the estimate for any single partial-survey year.

Apparently, the present method calculates $\hat{n}_i = m_i \frac{\sum_J N_j}{\sum_J m'_{ij}}$. This weighted mean would be appropriate under an assumption that the uncertainty in $\hat{n}_{ij}$ as an estimate of n_i is inversely proportional to m'_{ij}. This might apply if nesting chronology were relatively constant and the timing of partial surveys always optimal, but their duration very variable. If, on the other hand, the major source of the uncertainty were year-to-year variation in nesting chronology while the duration of partial surveys was relatively constant, this would be less appropriate, and the weighted mean would be biased upward. However, the uncertainty of the weighted mean, if used, would still be appropriately calculated from the variation between the $\hat{n}_{ij}$.

The present method of calculating the *uncertainty* of the $\hat{n}_i$ by combining activity records from all the saturation surveys and bootstrapping the entire data set, treating turtles as observational units, would be appropriate if turtles behaved independently and between-year differences in nesting chronology were due only to the chance *independent* arrival in one year of turtles with ideas different from those of the turtles that arrived in another year. This is unlikely to be the case—animals seldom behave independently—and the high precision reported for partial-survey estimates of numbers is probably spurious.

The same method would be applied *mutatis mutandis*, using the s_i instead of the m_i, to the estimation of nesting numbers from the number of individual turtles identified in the partial survey.

There seems to be some enthusiasm for calling this a Horvitz-Thompson estimate (the words 'Horvitz-Thompson' recur no less than 11 times in a single publication, including 3 times in a single figure caption—however great the need to impress, surely this is overkill?). The Horvitz-Thompson estimator relates to the problem of designing a procedure for sampling from a population with unequal probabilities without replacement—something rather different from this. This is an ordinary sample-survey estimate of numbers, with the feature that the sampling fraction is unknown and has to be estimated.

A major caveat on the suitability of the methods relates to the practice of counting only on East Island. It transpired in the meeting that in the course of the 30-year series, there have been changes, possibly significant, in the nesting habitat elsewhere in French Frigate Shoals and perhaps also elsewhere in the archipelago. Did we hear that entire islands have disappeared? However, journal publications refer to the rate of increase of the East Island numbers as though it applied quite certainly to the entire population, without mention of these alterations; to my mind somewhat dangerous, in that the articles could be challenged by a critic who might say that since the authors have not mentioned these changes they may be uninformed and their estimates of trend therefore unreliable. If the articles discussed the habitat changes, made reasonable estimates of their possible effects, and proposed some bounds on the resulting uncertainty of extrapolating the East Island trend to the entire archipelago, it would lead to a more defensible conclusion.

4.3 An evaluation of the experimental design and data collection methods for the marine habitat research. Are the data being collected in such a way that vital population parameters/rates such as: survival, recruitment, and relative abundance can be estimated?

No description of the data collection methods, and certainly not of the experimental design, was readily forthcoming; even direct questions on the experimental design elicited little information. As far as I could make out, the experimental methods and study design are not directed towards collecting population-level statistics, because the intention is to recapture tagged turtles ('You wouldn't want to go where there weren't any tagged turtles.') and collect longitudinal data on individual animals. This is not a criticism, as this objective is a valid one, is being attained, and may have been the original goal of the programme; it could be that the goal of 'population parameters' is a recent arrival.

However, a preliminary conclusion would be that the data are not being collected in such a way that *any* population vital rates can be estimated. This review cannot give a more definite answer to this question. This could be dealt with by a *focused* workshop dealing with no other subject but the at-sea sampling, its attainments, its original, present and possible future objectives, and the designs and methods that would be appropriate to them; and with a precondition that a detailed document fully describing the spatial and temporal intensity of both the capture effort and the captures—e.g. annual maps of both where effort was applied and the locations of captures, plus tabulations of the numbers of captures and recaptures, capture intervals, &etc—was prepared in advance.

It should however be noted that estimating any of the vital statistics mentioned—recruitment (understood as the rate at which new turtles come from the pelagic life-stage into the coastal pastures), survival, or relative numbers—through at-sea mark-recapture methods is likely to be a protracted undertaking, becoming yet more protracted as the population increases in numbers. The mathematics of mark-recapture is unkind. The precision of estimates depends on the number of recaptures, and if capture effort is distributed to make captures and recaptures independent in a large population, the recapture rate, and the precision of estimates, can be disappointingly small. Concentrating tagging effort spatially to ensure high recapture rates is efficient for getting longitudinal data on a restricted number of individuals, but invalidates population-level estimation.

If there is a coastal pasture in which capture and tagging effort is already well distributed and where the hypothesis that capture and recapture probabilities are independent can be defended, it could be appropriate to try a pilot analysis of its tag data as a mark-recapture data set, to see whether results can be obtained. It would, however, be necessary to carefully scrutinise the data, including the distribution of capture effort, as a component of the analysis, and if possible to include tests for independence of capture and recapture probabilities.

4.4 An evaluation of the analytical techniques used for trend analyses, somatic growth rates, and stock assessments with emphasis on Bayesian state-space modelling. Where necessary, the reviewers should recommend new or alternative analytical techniques.

In general, the documents reviewed show a tendency to use non-parametric multivariate models without exploring simpler alternatives first. The models used are described only in the most general terms amid clouds of nebulous irrelevancies. No equations are used to describe models and quantitative results are not tabulated; many (most) results are graphed on scales that are cryptic, incomprehensible, and not explained. Criteria for model fit are not well stated or documented.

Where Bayesian methods are used, there is poor, or no, exposition of the prior distributions used, and no documentation that the priors have been updated.

I don't know if it is at this point appropriate to recommend new or alternative analytical techniques—the whole analytical effort seems at the moment driven by a passion for the unfamiliar. I would suggest much more thorough exploration of parametric models before using non-parametric ones, which have the failing—as they are used here—that they give little quantitative output. What I would most strongly recommend would be 'new or alternative' writing techniques with a lot more effort seriously devoted to explaining what is being done and to cutting out the verbiage with which all these articles are so liberally padded, and less to trying to make it appear as complex as possible.

4.4.1 Trend analysis.

Qualitatively, the methods appear to be reasonable, and extract a trend and a cyclical component from the time series. However, these qualitative results are available from simple inspection of the data. At the quantitative level, the amplitude of the cyclical component is underestimated by a factor of close to 2, while the precision of the estimate of the rate of increase is overstated by a factor of about 4 (see below 6.1 and Appendix I). These errors invalidate the choice of method. A simple parametric model was easily fitted (see Appendix I) and gave more informative results—notably, that the period of the cycle is very tightly defined by the data, its amplitude much less so. In this case, the immediate flight to complex methods, capable of giving only qualitative results, instead of exploring the obvious, simple, quantitative, parametric model produced poor results.

4.4.2 Growth rates.

The growth-rate documents suffer from the same failing as the others, an inadequate presentation of models and methods. The methods might be appropriate, but give a strong impression of complexity unnecessarily multiplied. The results look plausible.

The fibropapilloma–growth article makes much mention of Bayesian methods, without defining the model, describing or justifying the priors used, or demonstrating that the priors are updated by the model fitting process. This induces caution in accepting these results at face value.

Converting pairs of size measurements to a growth estimate, then using a general model that includes size to explain growth, and integrating the resulting growth-size curve to arrive at a size-time curve is not very satisfactory; it would be better to fit size data directly to a size-time model, simultaneously including other independent variables such as severity of disease or home pasture.

4.4.3 Strandings.

The stranding analysis considers the distribution of strandings between causes, instead of considering separately the rate of stranding due to each cause. (There is some doubt about this, because the results graph something called ‘prob. xx stranding, centred scale’ with units that make no sense at all as a probability, so what is being presented is unclear).

The analyses are confined to the turtle stranding counts; stranding rates are not normalised even on an index of turtle numbers, let alone any consideration of the intensity of causative factors such as fishing effort in various fisheries, the number of boats likely to be involved in boat strikes, &etc.

This is a very confused and confusing article, referring endlessly to 'probabilities' instead of considering what ought to make more sense, stranding *rates*. The repeated use of 'mortality' to refer to the proportion of strandings that are dead when found is also a concern; 'mortality' is something else, and should not be confused.

4.4.4 Stock-dynamic assessment model with Bayesian fitting.

The data available and presented here will not support an assessment model of the type attempted. The results of the Bayesian fitting show simply that the data has failed to update the prior distributions of the parameters. The modelled population trajectories under exploitation depend not on the data, but on the median values of the input priors, and these are inadequately explained.

I suggest taking a step back and considering stock-structure models to generate a meaningful prior for a scaling factor between nester numbers and stock biomass, using only a logistic model, placing much less emphasis on estimating carrying capacity, and recognising that the present nester series shows no sign of a density-modulated reduction in rate of increase.

It might be appropriate to consider a simple stock-structure model, which could put bounds on the ratio between nester numbers and stock biomass. Given the documented increase in nester numbers of nearly 6% a year, this might indicate the present annual increase in biomass of harvestable stock. Some proportion of that increase might then constitute an allowable take. Bayesian methods might (or might not) be appropriate for combining prior distributions of stock-structure parameters, based on existing knowledge, to generate a posterior distribution of the nester: biomass ratio, which combined with the distribution of the rate of increase could generate a posterior for an allowable take.

But the present state of the data will not support a full-blown general-format Pella-Tomlinson assessment model, and something more modest would be more appropriate.

5. Reviews of Documents.

5.1 Trend

Balazs, G.H. and M. Chaloupka. In press. Recovery trend over 32 years at the Hawaiian green turtle rookery of French Frigate Shoals. Atoll Research Bulletin 00: 000–000.

Balazs, G.H. and M. Chaloupka. 2004 Thirty-year recovery trend in the once-depleted Hawaiian green sea turtle stock. Biol. Cons. 117: 491–498.

There is a tendency to consider the East Island nesting ground to be the same as the population—this is more marked in the Biol. Cons. article. As I have observed elsewhere, it transpired at the meeting that there have been changes in nesting beaches elsewhere even in French Frigate Shoals alone, let alone the rest of the archipelago, and to have included a discussion of the possible effects of habitat changes and of possible nesting elsewhere would have resulted in a more defensible and scientifically sounder product.

It is observed in the Biol. Cons. article—without reference or explanation—that ‘this constant level (scil. 32% each year) of apparent new nester recruits suggests that the . . . population might be approaching carrying capacity.’ This is difficult to understand. A population growing at a constant rate with stable demographic structure might be far from carrying capacity, but still adding new breeders at a constant rate; or a population that was stationary at carrying capacity might be turning over breeders at some other constant (but presumably lower) rate. It is in fact precisely when a population is *approaching* carrying capacity, the rate of increase is slowing down, and the demographics are not stable, but changing, that the proportion of new nesters might be expected *not* to be constant. It appeared at the meeting that this statement could not be defended. It is in my view not advisable to make casual statements about stocks with a history of over-exploitation that they are 'approaching carrying capacity' without some justification. The time trend of East I. nester numbers gives no indication that carrying capacity is being approached—in fact, it has been analysed as a constant increase, i.e. far even from its point of inflection, let alone carrying capacity.

There is also a statement that there is a 'strong cross-correlation between sea surface temperature and the STL annual nester remainder'—no correlation coefficient is given, and when I asked about it, I was told that there is no correlation; so where this statement is derived from remains slightly mysterious.

As far as the quantitative analysis is concerned the methods used were described as: 'a generalised additive model and Bayesian inference to account for any nonlinear trend and the uncertainty in the trend given the substantial interannual fluctuations in observed nester abundance, fitted using BayesX with random walk smoothness priors and a Bayesian smoothing spline (Fahrmeir and Lang, 2001)', as well as 'a procedure known as Seasonal and Trend decomposition using Loess or STL (Cleveland et al., 1990), which decomposes a series using nonparametric smoothing into additive frequency components of variation—(1) trend, (2) cyclical or quasi-periodic, (3) seasonal (if applicable using for instance a monthly data series) and (4) the residual or remainder.'

There is no further description of any model, the priors for the Bayesian components, the fitting criteria, or the degrees of freedom held in the splining processes, and what the independent variables were in any generalised additive model is never explained. As far as one can see, there is only one independent variable, and that is time, so where the 'additive' comes in is unclear.

The results obtained using these methods are displayed graphically in panels a–d of Fig. 3 of the Biol. Cons. article. *Qualitatively*, panels b and c are unexceptionable: there is an increasing trend, and there is a cyclical component. But this much is already evident from panel a, so the decomposition is no great gain. *Quantitatively*, the values graphed in panel c are too small by a factor of about 2 (see Appendix I; also compare the amplitude of the swings graphed in panel a with those graphed in panel c). Either the STL method is inherently faulty, or it is being wrongly applied.

The Atoll Research Bulletin article names a 5.7%/yr rate of increase with a 95% CI of 5.3–6.1, which converts to an e.c.v. of about 3.6%. This overstates the precision with which the trend can be estimated by a factor of about 4 (see Appendix I).

5.2 Growth analyses

Balazs, G.H. and M. Chaloupka. 2004. Spatial and temporal variability in somatic growth of green sea turtles (*Chelonia mydas*) resident in the Hawaiian Archipelago. Mar. Biol. 145: 1043–1059

Chaloupka M., and G. Balazs. 2005. Modelling the effect of fibropapilloma disease on the somatic growth dynamics of Hawaiian green sea turtles. Mar. Biol. 147: 1251–1260.

Both these articles use generalised additive models, not further described. About all one can say of either is that the results may well be qualitatively correct—on the available information, it is impossible to tell. The quantitative results are not easy to decode. The second of these articles makes considerable mention of Bayesian methods, with no description of the model used, no definition of the priors, and no demonstration that the priors are updated by the modelling. Given the disregard of elementary precautions in interpreting Bayesian results demonstrated in the assessment modelling article, there may be some difficulty in accepting these results at face value.

5.3 Strandings

Chaloupka, M., G.H. Balazs, S.K.K. Murakawa, R. Morris and T.M. Work. In prep. Cause-specific temporal and spatial trends in green sea turtle strandings in the Hawaiian archipelago (1982–2003).

This is an extremely confusing article, referring continually to 'probabilities'. The author even confuses himself in the caption to Fig. 3: it is panels b and d that present the same data, merely scaled on the overall proportion dead, not panels a and d. It is never clear whether what is being referred to is the rate of stranding assigned to a particular type, or the probability that, given that a stranding has occurred, it is of a particular type. It would be more useful simply to consider strandings of different types as being independent, and simply deal with them one at a time. The article is restricted in its approach, confining itself to the stranding data—there is no description of coastal rod-and-line fishing activity, nor of the gillnet or trawl fisheries operating in the archipelago, or of the trend in the number of recreational boats. Stranding rates are not normalised on the index of nester numbers or on any index of hazard intensity. There is a strong tendency to confuse proportions dead when found with mortality.

Captions of Figs 6–9 refer repeatedly to effects of each independent variable 'conditioned on' the other 3 variables. However, it transpired at the review meeting that these are not in fact the partial effects of the independent variables *after* the others have been fitted, but simply the marginal effects in a multivariate model.

5.4 Bayesian state-space assessment modelling.

Chaloupka, M. and G. Balazs. In prep. Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock. Unpub. Script. 38 pp.

Bayesian statistics is not just another method of fitting models to data. It is also uniquely dangerous, because it can pretend to produce results—estimates of model parameters—from data that inherently does not contain information. The present document is an example.

This document contains some equations defining the model used. Eq. 4 is fine; eq. 5 is wrong. Eqq. 6 and 9 are at best unconventional. The extension of the model to the so-called ‘Allee effect’ defined by eq. 11 defies logic. According to the conventional wisdom, and indeed to the text at the bottom of p. 17 of the present document, the Allee effect operates through reduced mating success and therefore reduced reproductive rates at low numbers. Eq. 11 leaves reproduction—the second term on the r.h.s. of eq. 11 (which incidentally still contains the same error as eq. 5)—unchanged, but introduces a large *mortality of the standing stock* at low numbers. This is inconsistent with the document’s description of an Allee effect. The trend diagram of Fig. 3 in the Biol. Cons. article on the increase of the stock shows that, at the start of the series when numbers were low, there was a period with a lower rate of increase. But to proclaim the existence of an ‘Allee effect’ on such slight evidence is hardly science—especially when the index of numbers being used is nesting females at a main breeding beach, and when there is in fact no suggestion that males are not being able to find their way to the same breeding location and successfully mate available females.

Returning to the results of running the Bayesian assessment model; in interpreting the results of Bayesian modelling, a usual (and essential) first step is to verify that the results do have some meaning, usually by documenting that the priors have been updated, and often by a cross-validation process verifying the prediction, by the model, of data values. Bayesian methods are dangerously willing to produce results where none exist. In the present case, there are clear deficiencies in the data. There is no information in the data that can scale the series of reported catches to the real world, as the reporting level is unknown, or to scale the nester index series to population size, or to scale the one series to the other. The population cannot be assumed at carrying capacity at the start of the catch series. The nester index shows no signs of an approach to carrying capacity, or indeed of any density-induced reduction of the rate of increase—in fact it has been modelled in an earlier article as showing a constant rate of increase implying that it is well below its point of maximum yield, and *a fortiori* nowhere near carrying capacity. With all these limitations in the data, it would be appropriate carefully to audit the results, checking (as an absolute minimum) that the priors have been significantly updated and that correlations between parameter estimates are not large.

The results tabulated are full of red flags. For variable after variable, the posterior confidence limits tabulated in Table 2 are the same as those of the priors tabulated in Table 1. In fact, for the carrying capacity K, the posterior is *wider* than the prior. For the Pella-Tomlinson shape parameter *z* it is obvious that the posterior, with a median at 2.98, 2.85, 2.76 and confidence limits near 1.08 and 4.89 is simply the unupdated prior, which was uniform from 1 to 5 having therefore confidence limits at 1.1 and 4.9. One would expect to see some recognition that the prior has not been updated—instead, there is a confident statement that:

'a skew-asymmetric function suggests that the compensatory process for the Hawaiian green turtle stock dynamics is a nonlinear function of density rather than the linear function implied by a logistic model (Fowler 1981). The 0.63K maximum productivity *estimate* for the Hawaiian green turtle stock is similar to productivity estimates for other large long-lived marine species such as pinnipeds and cetaceans.'

This is fiction. The ’skew-asymmetric’ function is the result of the prior, and nothing in the data has had anything to do with it. And the 0.63 K is in no sense an estimate, it is a function of the prior.

Despite the fact that the Bayesian modelling has produced no results, the ‘posterior’ median values are recklessly inserted into an exploitation model. The modelled trajectories under exploitation depend on 4 parameters: the carrying capacity, the rate of increase, the Pella-Tomlinson shape parameter, and the present state of the stock relative to carrying capacity.

For carrying capacity, the prior is not updated by the data, the posterior distribution is the unchanged prior (for some models, the posterior is *wider* than the prior). There is no information on how the prior was derived, merely a 2-line reference to some unpublished stock-structure model or other. Stock structure models do not often yield information on carrying capacity. A prior on carrying capacity might be obtained by considering the area of the coastal pastures, the productivity of the range, and the energetic requirements of the species, but would obviously need careful justification.

For the P-T shape parameter the prior is not updated; the posterior is the unchanged prior, uniform from 1 to 5. There is no statement of how this prior was derived. (The text at the top of p. 9 that this parameter is significantly greater than 1—simply because it is bounded by its prior—is remarkable.)

For the rate of increase, there is indeed information in the data. That we’ve seen before.

The present state of the stock relative to carrying capacity depends on the scaling factor between nester numbers and stock biomass; again, the prior is not significantly updated. The derivation of the prior is not explained; although here *is* an opportunity where a reasonable demographic model, with a spectrum of possible values for stock-structure parameters, might generate a legitimate informative prior.

All in all, the paragraph in this document reading:

‘Despite data limitations and some imprecise parameter estimates , the Bayesian state space surplus-production models nonetheless provided meaningful estimates of stock status and trend as well as some important population and management measures for the Hawaiian green turtle stock (Table 2, Figs. 2 and 3). We anticipate that these models can be used to assess the current recovery status of the Hawaiian stock (Fig. 2) and to determine whether a limited harvest for indigenous cultural purposes might be demographically feasible (Fig. 7).’

is alarming in the extreme; the Bayesian model produced *no* results on stock status (except to confirm the increasing trend in nester numbers) or management measures. This glib statement arouses grave misgivings about the usability of results presented in other documents reviewed. About all we can say on the basis of the available data is what we could deduce from the nester series that, provided the nester series is a useful index, the stock has been doing something near 6%/yr for the last 30 years and shows no signs of being near its point of maximum production or carrying capacity.

Appendix I.

Analysis of the series of estimated numbers of nesting females on East Island 1973–2002.

From the Biol. Cons. article a 'de-cycled' data set was constructed as the difference between Fig. 3.a and Fig. 3.c. A linear regression against time was fitted to this 'de-cycled' series. The result was a rate of increase of 5.7%/yr, with a st'd error of 0.85%, i.e. an e.c.v. of 15%. The residuals from this regression were strongly correlated with the fitted cyclical component of Fig. 3.c (corr. coeff't 0.506***) (the residuals graphed in Fig. 3.d are even more strongly correlated with Fig. 3.c (corr. coeff't 0.516)). The extraction of the cyclical component could therefore be improved. The correlation was set to zero if the cyclical effect of Fig. 3.c was multiplied by 2.07; this also minimised the total sum of squares about the model (i.e. optimised the fit) to 3.38 instead of 4.54, a reduction of some 25%. All this was consistent with the observation that the amplitude of the cyclical swings in Fig. 3.a of the Biol. Cons. article is roughly twice the amplitude of the extracted cyclical series in panel c. The methods used for decomposition in the Biol. Cons. article either are inappropriate, or are being wrongly applied.

For comparison, a simple parametric model was fitted to the data of Fig. 3a, i.e. the basic data. The model comprised a simple harmonic cycle (3 parms) and a linear trend (2 parms). It fitted very tidily, with a period of 3.86 years (s.e. 0.06 yrs) and a linear increase of 6.0%/yr (s.e. 0.83 %/yr). The sum of squares about the model was 4.49, i.e. slightly better than the published results, but worse than their optimised version. The small e.c.v. of the period (1.6%) indicates a cycle closely defined by the data, but the amplitude was poorly estimated with an e.c.v. of 21.6%, confirming the observation from Fig. 3.a and Fig. 3.c that the amplitude of the cyclic component is variable.

This simple model is susceptible of a simple description, can be written in one equation, is described by 5 parameters that can be easily understood and quantitatively reported (including a quantitative estimate of the period of the cycle and its s.e.), and took 15 minutes to fit using nothing more complicated than Microsoft Excel. The (half-) amplitude of the cyclical component was 0.47, which is consistent with doubling the cyclical component in the published article, with its amplitude of something near 0.2, to get the best fit. The more complicated non-parametric methods that have been used may have been more flexible in fitting the *variable* amplitude of the cyclical component, but have clearly been unsuccessful in fitting its *mean* amplitude.

The ARB article gives a 95% C.I. for the rate of increase of +/- 7% (+/- 0.4/5.7), corresponding to an e.c.v. of about 3.6%. It is not clear how this was obtained; possibly by running a linear regression though a smoothed trend(!). It constitutes, *prima facie*, a gross overstatement of the precision with which the increasing trend can reasonably be estimated; the data will support e.c.v.s of 14–15%.

Appendix II: Statement of Work

Consulting Agreement between the University of Miami and Dr. Michael Kingsley

STATEMENT OF WORK

CIE Review of Biological Data Collection Methods and Quantitative Analyses to Estimate Abundance and Trends for the Hawaiian Green Turtle Metapopulation

Background

The Hawaiian green sea turtle metapopulation has been increasing over the last 20+ years and appears to be recovering from overharvest prior to 1975. This population has been determined to be discrete in terms of molecular genetics and should be considered as a discrete stock. Obstacles to the continued wellbeing of this population may remain, including the impacts of sea level rise on nesting beach habitat, pollution, and other unforeseeable impacts such as disease. As such, it is important to know that the current monitoring and data analysis methods used for this population are adequate to detect changes in population trends at an early enough stage to take mitigation action. This will also be essential for consideration of delisting the threatened Hawaiian green sea turtle according to the US Endangered Species Act.

There are two primary components to the Hawaiian green turtle population monitoring project, nesting beach censuses and marine habitat mark-recapture. These components are supplemented with stranding, salvage and necropsy research to allow assessment of mortality sources and the relative risks.

Nesting beach green turtle research
Conducting counts on nesting beaches is the primary method used to monitor abundance trends in marine turtle population. Most of the nesting for this population occurs at French Frigate Shoals in the remote Northwestern Hawaiian Islands, making monitoring of the beaches logistically challenging. In addition, nesting green turtle habitat overlaps with Hawaiian monk seal pupping and nursing habitat and care must be taken not to disturb these endangered pinnipeds. Given these limitations, a monitoring regime that captures a segment of the nesting season in conjunction with statistical methods to estimate total nesting from the partial monitoring is used.

Marine habitat green turtle research
A recent petition letter from the NGO Oceana calls on NMFS to increase in-water abundance estimates for sea turtles and there was a focused discussion of this need at the recent NMFS National Marine Turtle Program meeting. Given that Hawaiian green turtles may take from 35-50 years to reach maturity, impacts to earlier life stages may not show up in nesting beach trends for decades, it is important to monitor trends in other life stages. The length of the in-water monitoring for the Hawaiian green turtle population is somewhat unprecedented and the techniques involved can potentially be considered a standard by which other monitoring projects

throughout the country can be set. Currently, data from the in-water portion of the study is primarily used for forage-site-specific somatic growth rates, though we are aware that such data, if collected appropriately, can be used for much broader purposes including survival rates, recruitment rates, and abundance trends.

Modeling efforts, data analysis

To date, several key papers have been published in peer review journals presenting analyses of trends in nesting females, forage site-specific differences in somatic growth rates, and fibropapilloma impacts on juvenile growth rates. In addition there is a manuscript currently under consideration for journal submission providing a stock assessment production model for the population.

Objectives of the CIE Review

The information presented for review has been developed by NOAA scientists and Dr. Milani Chaloupka under contract to NOAA Fisheries. The CIE review has been determined to be the best way to initiate and complete a thorough examination of the appropriateness of data collection methods and the analytical techniques used for Hawaiian green turtle recovery status and population monitoring. The information is to be examined by the CIE reviewers at a two-day workshop in Honolulu, Hawaii, on May 9-10, 2006, which will focus on analytical techniques and field methodologies used to acquire biological inputs.

Tentative Schedule of Presentations May 9-10, 2006

- Historical overview of green turtle research in the Hawaiian Island, (1972-2006) including past and present sampling methodologies – **George Balazs**
- Storage and retrieval of green turtle research data- Past and Present – **Shawn Murakawa**
- Green turtle stranding and salvage research and sample collection **– Cody Hooven**
- Diagnosis, treatment and captive-care rehabilitation of live-stranded green turtles in the Hawaiian Island **– Robert Morris, DVM**
- Health and disease assessments and necropsy research of green turtles with emphasis on fibropapilloma tumor disease in Hawaii **– Thierry Work, DVM**
- Development of quantitative techniques for monitoring and sampling numbers and trends of nesting green turtles at the Hawaiian green turtle rookery of French Frigate Shoals **– Jerry Wetherall**
- Genetics research of the nesting females and foraging immature green turtles in the Hawaiian Islands – **Peter Dutton**
- Modeling and other quantitative techniques for determining vital elements of the status and trends of the Hawaiian green turtle populations- with focus on journal publications and papers in progress **– Milani Chaloupka**

All presentations and sessions over the two-day period will be relatively informal and entirely interactive. During each presentation the two CIE reviewers will be encouraged to ask questions at any time. At the end of each presentation, additional time will be allotted for questions and discussion. At the end of the two-day period, specific time will be allotted for the reviewers to ask additional questions.

Specifically, the CIE reviewers shall analyze the reports on the Hawaiian green sea turtle metapopulation, focusing on the following:

1. An evaluation of the appropriateness of the quantitative methods being used for nesting beach research.
2. An evaluation of the experimental design and data collection methods for the marine habitat research. Are the data being collected in such a way that vital population parameters/rates such as: survival, recruitment, and relative abundance can be estimated?
3. An evaluation of the analytical techniques used for trend analyses, somatic growth rates, and stock assessments with emphasis on Bayesian state-space modeling. Where necessary, the reviewers should recommend new or alternative analytical techniques.

The CIE reviewers shall include the following reports in their analysis:

a) Balazs, G.H. and M. Chaloupka, 2004. Thirty-year recovery trend in the once depleted Hawaiian green sea turtle stock. Biological Conservation 117, 491-498.
b) Balazs, G.H. and M. Chaloupka. In Press. Recovery trend over 32 years at the Hawaiian green turtle rookery of French Frigate Shoals. Atoll Research Bulletin.
c) Chaloupka, M. and G. Balazs, 2005. Modelling the effect of fibropapilloma disease on the somatic growth dynamics of Hawaiian green sea turtles. Marine Biology 147, 1251-1260.
d) Chaloupka, M. Manuscript. Using Bayesian state-space modeling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock.
e) Chaloupka, M., G.H. Balazs, S.K.K. Murakawa, R. Morris, and T.M. Work. Manuscript. Cause-specific temporal and spatial trends in green sea turtle strandings in the Hawaiian Archipelago (1980-2003).
f) Wetherall, J.A., G.H. Balazs, and M.Y.Y. Yong. 1998. Statistical methods for green turtle nesting surveys in the Hawaiian Islands. *In* S.P. Epperly and J. Braun (comps.), Proceedings of the Seventeenth Annual Symposium on Sea Turtle Biology and Conservation, March 4-8, 1997, Orlando, Florida, p. 278-280. U.S. Dep. Commer. NOAA Tech. Memo. NMFS-SEFSC-415, 294 pp.

Reviewer requirements and schedule

Two reviewers shall be selected with the following expertise:

- One reviewer should have strong quantitative expertise, with some emphasis on Bayesian and stochastic modeling. Experience in designing statistical models appropriate to estimating total nesting females/nesting trends based on partial censuses, and the analysis of mark-recapture data for abundance trends and survival rates is essential.

- One reviewer should have an understanding of the difficulties inherent in marine turtle research and nesting beach monitoring, and experience with study designs for ocean habitat turtle sampling, including tagging methods, frequency sampling, and spatial considerations.

Each reviewer shall spend a maximum of 12 days for this review. The reviewers' duties shall include two days to participate in a workshop in Honolulu, Hawaii, on May 9-10, 2006, which will focus on analytical techniques and field methodologies used to acquire biological inputs. Additional days will be required for each reviewer to review documents prior to the meeting and to produce their individual written report of their findings subsequent to the workshop. No consensus, pre-final review, or rejoinder comments shall be required or will be accepted.

No later than May 24, 2006, the reviewer's reports shall be submitted to the CIE for review. See Annex 1 for details on the report structure. The CIE reports shall be sent to Dr. David Die, via e-mail to ddie@rsmas.miami.edu and to Mr. Manoj Shivlani via e-mail to mshivlani@rsmas.miami.edu.