On Statistical Methods that Permit and Improve the Use of Simpler Model-Based Strike Limit Algorithms

Geof H. Givens*

March 19, 1999

Abstract

The difference between the dynamics of a population model in an Aboriginal Whaling Management Procedure (AWMP) and the dynamics of either the true stock or the common control program for testing AWMPs is a form of model misspecification. The potential estimation bias introduced by this misspecification provides the opportunity to empirically improve model-based strike limits. We show how to adjust the application of an existing AWMP in order to exploit new data sources or to use old data sources more appropriately. The result can be better strike limits without the need to design ever more complex population dynamics models. For example, age proportion data can be used with the AWMP introduced by Punt and Butterworth (1997) without redesigning their model to be age-stratified. A simple example is presented to isolate the key aspects of the approach. For this example, the proposed approach provides better estimation through more flexible use of all available data, particularly when the informativeness of portions of the data differ.

INTRODUCTION

In its effort to develop Strike Limit Algorithms (SLAs) as part of an Aboriginal Whaling Management Procedure (AWMP), the Standing Working Group on the Development of an AWMP (SWG) has relied on input from scientists with a wide variety of expertises. One dimension along which their contributions may be classified is statistical empiricism versus careful biological modeling. For example, the idea of H-optimisation (Givens, 1997; 1999a) for improving SLAs by empirically fitting a Bayes rule tuning of a nominal SLA is a purely statistical tool that ignores biology. In contrast, the SLA introduced by Punt and

*Geof H. Givens is Assistant Professor of Statistics at the Department of Statistics, Colorado State University, Fort Collins, CO 80523.
Butterworth (1997) relies on a complex, biologically-motivated simulation model that might be questioned by believers of the statistical maxim that 'no model is correct'.

That maxim is meant to convey the idea that virtually all models of complex phenomena are misspecified in that they do not perfectly simulate observed processes, even though many such models are very useful, including the one used by Punt and Butterworth. In the context of Initial Exploration Trials, the Punt and Butterworth (1997) SLA is misspecified by design: the SLA uses a different population dynamics model than the one used by the SWG to simulate stocks and test SLA performance. Such misspecification is reassuring because it helps avoid over-reliance on the simulation testing model which only roughly reflects reality. Model misspecification can, however, degrade estimation performance.

Complex modeling such as in the Punt and Butterworth (1997) SLA includes a strong dose of statistical empiricism because such models are fit to some observed data. The quality of that fit depends on the degree to which the model is misspecified. For example, fitting a straight line to a scatterplot of quadratic data can lead to poor estimates in some circumstances and adequate ones other times.

A drawback to a pure modeling approach is that it can be difficult to incorporate new sources of data as they arise. For example, the dynamics model used by the Punt and Butterworth (1997) SLA is not age-stratified. If useful age-structure data are available for a stock, then a pure modeler must either argue that the new data are irrelevant or uninformative, or go back to the drawing board and design a more complex model (e.g. an age-stratified model) that incorporates parameters related to the new data. Depending on what model revisions are needed, this process of adding model complexity can be quite difficult and the more complex model is often more difficult to fit.

When a purely model-based AWMP is designed for application to multiple stocks, it must rely upon only those data that are available for all stocks. Therefore, for stocks for which more than than the minimal data are available, such an AWMP would ignore potentially useful information. An overly simple and misspecified model-based AWMP may be used for this reason or merely because it is the best that could be done with limited time and effort. In either case, it would be nice if there was a simple way to enhance the AWMP for stocks where more data are available or where one wishes to weaken the dependence of the AWMP on the particular model used.

For example, consider the fishery type 2 Initial Exploration Trials (IWC, 1999) and the Punt and Butterworth (1997) SLA. This SLA ignores the series of age structure data (observed counts of calves and mature animals six times in the past and every five years after AWMP management begins (Angliss et al., 1995; Givens et al., 1995). To revise this SLA so that it used an age-stratified population dynamics model and could therefore incorporate these age data into its likelihood would be a tedious job. Can we determine in advance whether that job must be undertaken so that the age data are used to improve strike limits? Is there any way to use all available data to improve strike limits while circumventing this massive model redesigning?

The answer is 'yes' to both these questions. Givens (1997; 1999a) has shown how an empirical, Bayes rule estimation approach (called H-optimisation by the SWG) can be used.
on a stock-specific basis to augment a model-based AWMP. This permits the use of available data and simple, existing AWMP models, rendering the design of a new, more complex model unnecessary. This hybrid approach, which draws on the strengths of both statisticians and biological modelers, is briefly reviewed in the next subsection. However, the purpose of this paper is to provide justification for the claim that this approach can produce AWMPs as good or superior to those that result from tediously designing ever more complex biological models.

Some attractions of this approach for mixing empirical statistical models with biological models are that:

- It is a faster way to develop, simulation test, and tune an AWMP.
- It allows the use of simpler biological models, even when such models are known to have deficiencies or to ignore sources of data.
- It can produce results at least as good as a purely model-based approach.
- It is motivated by a statistically optimal estimation strategy known as the Bayes rule.
- There's no reason to believe that adding complexity to a biological model must reduce the degree to which the model is misspecified, however the proposed approach can reduce the misfit.

Some drawbacks of this approach (with counterarguments) are that:

- It includes a component where a purely statistical, empirical fit is estimated. If one insists that models used by an AWMP must be wholly derived from biological principles, than the approach proposed here is unacceptable. Portions of the Revised Management Procedure for Commercial Whaling are clearly chosen for performance or convenience rather than for biological accuracy; thus insistence on pure modeling would be a higher standard than has been achieved in the past.

- It will not necessarily eliminate model misspecification. But it can reduce the misfit.

- It requires a developer to explicitly state relative probabilities or weights for some Initial Exploration Trials. However, this task falls to the AWMP developer, not the SWG, and hence choice of these weights becomes part of an individual's development process. Whatever choices the developer makes, the resultant AWMP will be exhaustively performance-tested by the SWG in a process that currently requires no such probabilities or weights. Thus, this requirement is akin to requiring that an AWMP developer select default values for the tuning parameters when s/he proposes an AWMP to the SWG.
Quick review of some H-optimisation ideas

Let $\theta$ denote the values of unknown parameters such as MSYR and $K$ which constitute the assumptions of a particular Initial Exploration Trial. Given $\theta$, let the sequence of idealized strike limits be denoted $H(\theta)$ where $H$ is defined, for example, as in Section E, Appendix 3 of IWC (1999). Throughout this paper, time and other nuisance subscripts are suppressed.

Suppose that one is using a biological model to represent how stock dynamics behave, given $\theta$. When relevant observed data, $D$, are used to fit the model, that usually means that $\theta$ is estimated by a value $\hat{\theta}(D)$ where $\hat{\theta}$ is chosen so that $D$ is very likely if $\hat{\theta}$ were the true parameter and the model exactly reflected the true biology and data generation process. Then the ideal strike limits, $H(\theta)$, are usually estimated as $Q(\hat{\theta}(D))$, which could be more simply written as $\hat{Q}(D)$. Note that this procedure relies on misspecified modeling since the model will not represent the truth completely accurately; neither must $Q$ and $H$ agree.

If a developer wishes to improve the performance of the SLA on a variety of trials to which s/he assigns weights $p(\theta)$, then an example of the H-optimisation approach would be (roughly) to minimize

$$\sum_{\text{time}} \sum_{\text{replicates}} \int (\alpha_0 + \alpha_1 \hat{Q}(D) - H(\theta))^2 L(D|\theta)p(\theta) \, d\theta$$

with respect to $\alpha_0$ and $\alpha_1$ where $L(D|\theta)$ is the likelihood function and replicate simulations of each trial are all run over a fixed time span. This minimizes the average posterior squared error between $\hat{Q}$ and the ideal $H$ over replicate Initial Exploration Trials. The resultant SLA would be $\hat{\alpha}_0 + \hat{\alpha}_1 \hat{Q}(D)$ where the $\hat{\alpha}_i$ solve the minimisation problem.

Several important details about this approach, such as how best to incorporate the notion of aboriginal need and how the resultant SLA can be considered an estimated Bayes rule solution to a compound decision problem, are provided by Givens (1997; 1999a).

Mixing biological and empirical models with H-optimisation

Now, consider again the motivating example of what to do with the Punt and Butterworth (1997) SLA if one is interested in using the age structure data. One could ignore the data and use the SLA as is (or a simple H-optimised version of it such as $\hat{\alpha}_0 + \hat{\alpha}_1 \hat{Q}(D)$).

Instead, one could try to replace the model used in the Punt and Butterworth (1997) SLA with a more complex, age-structured version so that a joint likelihood, say $L(D, A | \theta)$ where $A$ is some summary of the age data, could be used to estimate model parameters. If the resulting SLA were H-optimised, one would minimise, say,

$$\sum_{\text{time}} \sum_{\text{replicates}} \int (\alpha_0 + \alpha_1 \hat{Q}(D, A) - H(\theta))^2 L(D, A|\theta)p(\theta) \, d\theta$$

The resultant SLA is $\hat{\alpha}_0 + \hat{\alpha}_1 \hat{Q}(D, A)$ where the $\hat{\alpha}_i$ solve the minimisation problem. However, the Punt and Butterworth (1997) SLA is already very complex, requiring hours to run 100
replicate Initial Exploration Trials; the fancier model \( Q(D, A) \) needed to provide \( L(D, A \mid \theta) \) would not be easy to create or use. Moreover, there is no guarantee that such a model would be less misspecified. If the fancier model failed to produce better strike limits after incorporating the age data, one could not infer whether this signified the uninformativeness of the age data or merely reflected the model misspecification.

The final option would be to retain the simpler, non-age-structured model and to incorporate the age data directly using H-optimisation. For example, one could minimize

\[
\sum_{\text{time}} \sum_{\text{replicates}} \int (\beta_0 + \beta_1 \hat{Q}(D) + \beta_2 A - H(\theta))^2 L(D \mid \theta)p(\theta) \, d\theta \tag{3}
\]

with respect to the \( \beta_i \) where \( A \) represents a summary of the age data. The resultant SLA is \( \hat{\beta}_0 + \hat{\beta}_1 \hat{Q}(D) + \hat{\beta}_2 A \) where the \( \hat{\beta}_i \) solve the minimisation problem.

Exploratory examples of how this could work with actual SLAs have been shown elsewhere (Givens, 1997; 1999ab; Givens and Bernstein, 1998). The remainder of this paper is a simple, abstract example unrelated to whales. This example illustrates the notion of model misspecification and permits analytical solutions for the optimal estimators. This allows attention to focus on how and why the H-optimisation approach differs from the fully biological approach.

**EXAMPLE**

**Introductory Notation and Relationship to Motivating Example**

Suppose two independent sets of data are available for estimation of an unknown \( H(\theta) \), and that the datasets may be summarized by statistics \( T_1 \) and \( T_2 \), respectively. Consider \( T_1 \) to be analogous to abundance data and \( T_2 \) to age proportion data. With trial weights given by \( p(\theta) \) and with joint likelihood \( L(T_1, T_2 \mid \theta) \), the best estimator of \( H(\theta) \) under squared error loss, say \( \hat{E} \), is the posterior mean of \( H \), namely \( \int H(\theta)p(\theta \mid T_1, T_2) \, d\theta \), where \( p(\theta \mid T_1, T_2) \propto p(\theta)L(T_1, T_2 \mid \theta) \) is the posterior for \( \theta \) given all the data. In other words, \( \hat{E} = \int H(\theta)p(\theta \mid T_1, T_2) \, d\theta \) minimizes \( \int (E - H(\theta))^2 p(\theta \mid T_1, T_2) \, d\theta \) (Lehman, 1983).

In many situations, however, the class estimators one has available may not include \( \hat{E} \). Suppose that someone who expends the effort to develop a model using both sources of data creates the estimator \( \hat{Q}_{\text{full}}(T_1, T_2) \). This would be analogous to developing a more complex, age-structured version of the Punt and Butterworth (1997) SLA. Then, allowing a simple H-optimisation, this developer has available a family of estimators of the form \( E_{\text{full}} = \alpha_0 + \alpha_1 \hat{Q}_{\text{full}}(T_1, T_2) \) with tuning parameters \( \alpha_0 \) and \( \alpha_1 \). The tuning step is sensible if \( \hat{Q}_{\text{full}} \) is an estimator based on a misspecified model. The values of the tuning parameters are chosen to minimize posterior squared error loss (analogously to (2)), thus yielding the
Bayes rule within the given family of estimators. However, if there are no values of $\alpha_1$ and $\alpha_2$ which yield $E_{f_{alt}} = \hat{E}$, then $E_{full}$ is misspecified and there is room for improvement.

Suppose another developer foregoes the effort to develop a biological model using both data sources and instead relies on a simpler model $\hat{Q}_{alt}(T_1)$ that ignores $T_2$. This would be analogous to using the existing Punt and Butterworth (1997) SLA. Again, since $\hat{Q}_{alt}$ relies on a presumably misspecified model, one could consider a tuned version. However, there is no reason why $T_2$ (or even $T_1$ itself) cannot be used in the tuning as well. This leads to the class of estimators, for example, $E_{alt} = \beta_0 + \beta_1 \hat{Q}_{alt}(T_1) + \beta_2 T_2 + \beta_3 T_2^2 + \beta_4 T_1 T_2$. Now the tuning parameters $\beta_i$ ($i = 0, \ldots, 4$) can again be chosen to minimize posterior squared error loss (analogously to (3)).

Which approach is a better way to use the information in $T_2$? The answer is that it depends on several factors including the degree of model misspecification inherent in the two nominal estimators ($\hat{Q}_{full}$ and $\hat{Q}_{alt}$) and the relative informativeness of the two data sources. The goal of this paper is to show that the second approach can produce estimators at least as good as the first approach. Hence, if an adequate $\hat{Q}_{alt}$ is available but a suitable $\hat{Q}_{full}$ would require substantial effort to develop, a reasonable strategy would be to use $\hat{Q}_{alt}$. The effort saved by not developing $\hat{Q}_{full}$ can instead be focussed on optimizing the development and tuning of estimators based on $\hat{Q}_{alt}$.

Specifics

Consider the following example\(^1\). Let the data be distributed independently according to $T_1 \sim N(\theta, \sigma^2/n)$ and $T_2 \sim N(a \theta, \sigma^2/(\gamma^2 n))$. Here $a$ and $\gamma$ will be used to manipulate the bias and precision of $T_2$ relative to $T_1$. The use of $n$ is to help imagine that $T_1$ and $T_2$ are ordinary sample averages, and the use of $\gamma$ obviates the need to consider different sample sizes. Let the prior $p(\theta)$ be $\theta \sim N(0, \sigma^2/(cn))$. Here $c$ will be used to manipulate the strength of the prior relative to the likelihood.

Let $H(\theta) = \theta^2$. The example begins by supposing that a misspecified ‘biological model’ is used to develop an estimator, as is done in AWMP development. Suppose this erroneous approach (taken by both developers) is to square the average of all available data, i.e. $\hat{Q}_{full}(T_1, T_2) = ((T_1 + T_2)/2)^2$ and $\hat{Q}_{alt}(T_1) = T_1^2$. These are potentially reasonable

---

\(^1\)In this example, we will eliminate the nuisance time variable by assuming that a strike limit is needed at only one time, and we will employ what Berger (Berger, 1980) terms a ‘normal’ Bayes analysis which averages over all possible data. An ‘extensive’ approach averages over the posterior given the observed data. The ‘normal’ approach is used here because, being independent of random data, it is easy to graph. The ‘extensive’ approach given by (1) is recommended for real applications. A simpler way to view this approach is to suppose that one has an infinite supply of Initial Exploration Trial replicates.

Without these assumptions, the results from this example would depend on the particular data observed. The effect of these assumptions is merely to examine mean performance. In the figures which summarize this example, the surfaces of relative performance are therefore mean surfaces, about which random variation would occur if we did not average over the data.

These technical simplifications allow the summations in (2) and (3) can be replaced by integrals, and also allow exact analytical computation of all the results which follow.
estimators: for example a developer trying to estimate $\theta^2$ might erroneously rely on the model $T_1 \sim N(\theta^2, \sigma^2/n)$, in which case s/he would believe $\hat{Q}_{alt}$ to be optimal. Similarly, assuming a particular form of density dependence, or deterministic dynamics, or constant $K$ or MSYR, or a hundred other such choices will lead developers to creating good SLAs which are nevertheless misspecified.

Let $E_{full} = \alpha_0 + \alpha_1 \hat{Q}(T_1, T_2)$ and $E_{alt} = \beta_0 + \beta_1 \hat{Q}_{alt}(T_1) + \beta_2 T_2 + \beta_3 T_2^2 + \beta_4 T_1 T_2$. Note that no choice of the $\alpha_i$ yields $\hat{E}_{full} = \hat{E}$. The same is true for the $\beta_i$ and $\hat{E}_{alt}$. Both classes of estimators originate from misspecified models, and the best tuning of each will match $H(\theta)$ less well than $\hat{E}$ would have.

Then, it can be shown that

$$\hat{\alpha}_0 = b^2\sigma^2(1+\gamma^2)/(nb^2\gamma^2(1+a)^2 + \sigma^2(1+\gamma^2))$$
$$\hat{\alpha}_1 = 4n^2b^4\gamma^4(1+a)^2/(nb^2\gamma^2(1+a)^2 + \sigma^2(1+\gamma^2))^2$$

is the tuning the yields the best tuning, on average. This tuning defines $\hat{E}_{full}$, which minimizes the average posterior risk at a value of

$$R_{full} = \frac{4b^4\sigma^2(1+\gamma^2)(nb^2\gamma^2(1+a)^2 + \sigma^2(1+\gamma^2))}{(nb^2\gamma^2(1+a)^2 + \sigma^2(1+\gamma^2))^2}.$$  

Similarly, the best tuning of $E_{alt}$ on average is

$$\hat{\beta}_0 = b^2\sigma^2/(na^2b^2\gamma^2 + nb^2 + \sigma^2)$$
$$\hat{\beta}_1 = n^2b^4/(na^2b^2\gamma^2 + nb^2 + \sigma^2)^2$$
$$\hat{\beta}_2 = 0$$
$$\hat{\beta}_3 = n^2a^2b^4\gamma^4/(na^2b^2\gamma^2 + nb^2 + \sigma^2)^2$$
$$\hat{\beta}_4 = 2n^2ab^4\gamma^2/(na^2b^2\gamma^2 + nb^2 + \sigma^2)^2$$

which provides the minimal average posterior risk for $\hat{E}_{alt}$ of

$$R_{alt} = \frac{2b^4\sigma^2(2na^2b^2\gamma^2 + 2nb^2 + \sigma^2)}{(na^2b^2\gamma^2 + nb^2 + \sigma^2)^2}.$$  

To determine which estimator is superior in this example, we compare them using $S = R_{full}/R_{alt}$. When $S > 1$, the best tuning of the alternative estimator is superior to the best tuning of the full estimator.

Fig. 1 plots $S$ versus $a$ and $\gamma$ when the prior is relatively weak, specifically $c = .05$. In this case, we see that $S$ always exceeds 1, indicating superior performance for $\hat{E}_{alt}$. This improvement is greatest when $T_2$ contains poor information: either a very precise signal biased towards unlikely values of $\theta$, or a very imprecise signal (particularly when this signal is biased towards likely values of $\theta$ since then $\hat{E}_{alt}$ uses $T_2$ more wisely). The situations in which $\hat{E}_{full}$ and $\hat{E}_{alt}$ are roughly equivalent are problem-specific and depend in complex manner on the relative bias ($a$) and informativeness ($\gamma$) of the two data sources. It would not be reasonable to assume that these two estimators are usually equally good.
Fig. 2 plots $S$ versus $c$ and $\gamma$ when $T_2$ has unbiased information, i.e. $a = 1$. In this case, $\hat{E}_{alt}$ is again generally superior, with best relative performance occurring when $T_2$ is either much more or much less precise than $T_1$. The superiority of $\hat{E}_{alt}$ is also weakly increasing as uncertainty about the true value of $\theta$ increases.

Fig. 3 plots $S$ versus $c$ and $a$ when $T_1$ and $T_2$ are equally precise about $\theta$, i.e. $\gamma = 1$. Here, the two estimators perform more similarly, but again $\hat{E}_{alt}$ is usually slightly superior with the greatest gains occurring when $T_2$ is biased. The superiority of $\hat{E}_{alt}$ also increases slightly as uncertainty about the true value of $\theta$ increases.

Finally, Fig. 4 plots $S$ versus $c$ and $a$ when $T_2$ is half as precise about $\theta$ as $T_1$. In this case, $\hat{E}_{alt}$ is again generally superior. There is a notable positive interaction which produces strongly superior performance for $\hat{E}_{alt}$ when $T_2$ is biased towards likely values of $\theta$ and there is a lot of prior uncertainty about $\theta$.

Together, these results show:

- $\hat{E}_{alt}$ can provide equivalent or superior estimation even though it does not use the $T_2$ data ‘properly’ by incorporating it in a biological model; and

- $\hat{E}_{alt}$ is a superior estimator in this example when
  
  - the influence of $T_2$ on estimated strike limits should be decreased because $T_2$ provides biased information about true stock status; or
  
  - the relative influences of $T_1$ and $T_2$ on estimated strike limits should be unequal because $T_2$ is either more or less informative about true stock status than is $T_1$.

CONCLUSIONS

The alternative, H-optimisation approach for mixing biological models with purely empirical components works because it is more flexible. The more complex, purely biological model is constrained to use the data in a particular way. If the model is misspecified, the data are misused. The H-optimisation approach allows a more adaptable approach to extracting information from the data.

Consider expanding $\hat{E}$ in a Taylor series with respect to $T_1$ and $T_2$. Write this expansion as $\hat{E} = c_0 + c_1 T_1 + c_2 T_2 + c_3 T_1 T_2 + c_4 T_1^2 + c_5 T_2^2 + c_6 T_1 T_2 + \ldots$. Then, the family of estimators $\alpha_0 + \alpha_1 \hat{Q}_{alt}(T_1) + \beta_2 T_2 + \beta_3 T_2^2 + \beta_4 T_1 T_2$ has tuning parameters which can be used to individually match several different powers of $T_1$ and $T_2$. Clearly the second family has a more flexible tuning parameterization that might more easily match dominant Taylor series terms. Hence it should not be surprising if the alternative approach produces an adequate or superior estimator. However, the $\beta_i$ are
Figure 1: Plot of $S = R_{full}/R_{alt}$ as a function of $a$ and $\gamma$ when $c = .05$. When $S > 1$, the performance of $\hat{E}_{alt}$ is superior to that of $\hat{E}_{full}$.

Figure 2: Plot of $S = R_{full}/R_{alt}$ as a function of $c$ and $\gamma$ when $a = 1$. When $S > 1$, the performance of $\hat{E}_{alt}$ is superior to that of $\hat{E}_{full}$. 

9
Figure 3: Plot of $S = R_{full}/R_{alt}$ as a function of $c$ and $a$ when $\gamma = 1$. When $S > 1$, the performance of $\hat{E}_{alt}$ is superior to that of $\hat{E}_{full}$.

Figure 4: Plot of $S = R_{full}/R_{alt}$ as a function of $c$ and $a$ when $\gamma = 0.50$. When $S > 1$, the performance of $\hat{E}_{alt}$ is superior to that of $\hat{E}_{full}$. 
not chosen using a strategy that aims to match \( \hat{E} \), so the alternative approach does not guarantee a superior estimator.

Although H-optimisation does not require the use of any biological modeling, it would be a mistake to eschew biological models altogether. At least some biological modeling is essential for an interpretable SLA that can be checked for scientific sensibility. Furthermore, biological modelers may have insights about the relationship between the data and appropriate strike limits that H-optimisers fail to adequately consider among their classes of relatively simple polynomial empirical models of the sort suggested in this paper.

Although biological models have been repeatedly referred to as 'misspecified' in this paper, it should be stressed that this does not mean that they fail to provide good strike limits. The polynomial empirical models employed by H-optimisation are also misspecified. Both strategies have positive contributions to offer AWMP development.

ACKNOWLEDGEMENTS

This research was partially supported by Contract No. 99-181 from the North Slope Borough (Alaska), the State of Alaska (through the Alaska Department of Community and Regional Affairs), and the National Oceanic and Atmospheric Administration (through the National Marine Mammal Laboratory to the Alaska Eskimo Whaling Commission), Purchase Order #40ENNF900079 from the National Marine Fisheries Service and Purchase Order #40ABNF801829 from the National Marine Mammal Laboratory.

REFERENCES


