MINISTRY OF FISHERIES
Te Tautiaki i nga tini a Tangaroa

ISSN 1175-1584

Preliminary evaluation of maintenance management procedures for New Zealand rock lobster (*Jasus edwardsii*) fisheries

Paul A. Breen
Susan W. Kim
Nokome Bentley
Paul Starr

New Zealand Fisheries Assessment Report 2003/20
May 2003

Preliminary evaluation of maintenance management procedures for New Zealand rock lobster (*Jasus edwardsii*) fisheries

Paul A. Breen[1]
Susan W. Kim[1]
Nokome Bentley[2]
Paul Starr[3]

[1]NIWA
Private Bag 14901
Wellington

[2]Trophia
PO Box 60
Kaikoura

[3]61A Rhine Street
Island Bay
Wellington

New Zealand Fisheries Assessment Report 2003/20
May 2003

Published by Ministry of Fisheries
Wellington
2003

ISSN 1175-1584

Citation:
Breen, P.A.; Kim, S.W.; Bentley, N.; Starr, P.J. (2003).
Preliminary evaluation of maintenance management procedures for
New Zealand rock lobster (*Jasus edwardsii*) fisheries.
New Zealand Fisheries Assessment Report 2003/20. 65 p.

This series continues the informal
New Zealand Fisheries Assessment Research Document series
which ceased at the end of 1999.

EXECUTIVE SUMMARY

Breen, P.A.; Kim, S.W.; Bentley, N.; Starr, P.J. (2003). Preliminary evaluation of maintenance management procedures for New Zealand rock lobster (*Jasus edwardsii*) fisheries.

New Zealand Fisheries Assessment Report 2003/20. 65 p.

This study explores "new decision rules" for managing rock lobster stocks. We report on a literature search and propose that the term "management procedures" be used to describe fully developed and tested procedures such as the current rule for the NSS substock, and that "decision rule" be limited to rules that have not been tested or that do not specify what action is to be taken upon triggering.

This document describes the various harvest control rules we encountered in the literature search, and presents additional rules we invented for this study. It describes a complete evaluation system, comprising population, observation, harvest control rule and implementation models. A set of fishery indicators is defined. All the harvest control rules described were tested informally, some were rejected and a subset was tuned for further testing.

Both base case results and the results of robustness trials are reported. Robustness trials involved changes to the underlying population dynamics, changes to the implementation model and changes to the observation model.

Rules performed similarly except for one rule with a much higher crash rate than the others. Robustness trials were more informative than the base case trials, confirming that finding a robust rule is more critical than finding an optimum rule. Most rules were sufficiently robust to the realistic robustness trials, but no rule could deal with uncontrolled catch when that catch was either too large or was allowed to increase. This is a serious shortcoming to the use of management procedures: it suggests that management procedures cannot be used unless the recreational, traditional and illegal catches are constrained by the management strategy.

CONTENTS

1. INTRODUCTION

In this paper we discuss work to develop a maintenance "decision rule" or "management procedure" designed to maintain the stock at or near some target level.

Existing decision rules in New Zealand rock lobster fisheries are of two types:

- the rebuilding rules (1997 and 2002 versions) used in the NSS stock, designed through extensive simulation testing to cause biomass to reach a target within some reasonable period, and
- for the NSN and NSC stocks, simple comparison of CPUE with a base level, triggering a stock assessment when CPUE falls below the base level.

The major differences between these two types of rules are the amount of testing that has been done and the degree to which the rule specifies what will be done based on the result. Neither the 2002 NSS rule nor the extant NSN/NSC rule is necessarily suitable for maintaining biomass at or near an optimum level. The NSN/NSC rule triggers nothing until CPUE is low, then triggers a stock assessment. The NSS rule is designed to rebuild the stock from the current low level to a higher level, but its performance once biomass has reached the target has not been evaluated.

The study described here was done under Objective 7 of the Ministry of Fisheries contract CRA 2000-01, which is to "evaluate new decision rules".

Our approach to this objective follows recent work in designing and evaluating the 2002 NSS "decision rule" (Bentley et al. unpublished results) or "management procedure" (in this study we will try to distinguish these terms; see Section 3) and follows current worldwide usage. Good reviews were provided by Butterworth & Punt (1999) and McAllister et al. (1999). These authors described using a system of sub-models, including at least a population simulator or "operating model", an observation model that simulates the population signal, and the harvest control rule model. Catch determined from the harvest control rule model is fed back into the population model in a feedback loop, to make a single run of 20 – 100 years. This whole process is repeated with simulated stochastic error, discussed further below, to make a large set of runs from which the distributions of indicator values can be examined. In turn, that whole process can be repeated for different variants of a specific harvest control rule, for completely different rules, for variants of the population model, or for models with different basic structure, and for different simulated realities.

This approach has already been used to evaluate the 1997 and 2002 NSS management procedures (Starr et al. 1997; Bentley et al. unpublished results), and management procedures for Hooker sea lion bycatch in the Auckland Islands arrow squid (*Nototodarus sloanii*) fishery (Breen et al. 2003).

In this document we report on a literature search conducted to see what approaches have been taken elsewhere and what harvest control rules have been developed. We describe the conclusions others have drawn from that work. We describe our system of sub-models and some tuning of the population model to data and previous results from the NSN stock. We describe a list of candidate harvest control rule types and report on preliminary evaluations, report more detailed evaluations from a subset of these candidates, and report on a series of robustness trials.

2. LITERATURE SEARCH

2.1 Management procedures

Evaluation of management strategies began in the late 1970s (Hilborn 1979) and has been variously called, for instance, management strategy evaluation (MSE), decision analysis framework, simulated adaptive control analysis, and harvest strategy evaluation. The recent literature suggests adoption of the standard term "management procedure" (Butterworth et al. 1997) for the results of this work. The management procedure approach was developed in South Africa (Butterworth et al. 1997, Cochrane et al. 1998), has been adopted by the International Whaling Commission (IWC) (Kirkwood 1997), and has spread widely (see reviews by Butterworth & Punt 1999, McAllister et al. 1999).

A management procedure is "a fully specified feedback control system applied as part of a fishery management system" (McAllister et al. 1999) and specifies:

- what data will be collected,
- how they will be collected and processed,
- what estimates will be made from the data, and
- how those estimates will be used to set harvest controls.

The advantages of this approach over the traditional pattern of regular or periodic stock assessments, each followed by a decision process, are (loosely based on Geromont et al. 1999):

- uncertainty in all facets of the assessment and management process can be addressed,
- harvest control rules can be developed that are robust to uncertainty,
- the process leads to explicit definition of management objectives,
- all participants in the fishery can become involved in the choice of rule,
- a long-term view is forced,
- management procedures move away from regular assessments, freeing resources for other research, and
- the process is more understandable to fishers than the traditional approach.

Robb & Pertman (1998) stressed the need for stakeholders to agree on objectives. With a management procedure comes a necessity for politicians and government managers to move away from "tactical" thinking ("what should next year's TAC [total allowable catch] be?") to "strategic" thinking ("what should the harvest control rule be?") (Butterworth & Punt 1999). These authors also suggested that management procedures tend to shift the attention of interested parties away from catch levels onto the data that are used by the management procedure, such as CPUE.

Components of the feedback loop (McAllister et al. 1999) are (Figure 1):

- a population model,
- an observation model,
- a stock assessment model (if the procedure is model-based),
- a harvest control model,
- a harvest decision model (often combined with the rule model), and
- an implementation model (important but often omitted)

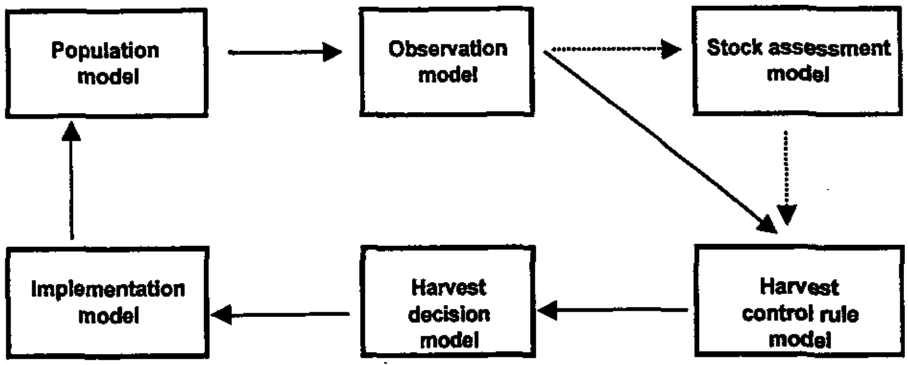

Figure 1: Main components in a quantitative fishery management system evaluation (from McAllister et al. 1999).

If the management procedure is model-based, then the stock assessment process must be simulated by the stock assessment model, perhaps for each year of each evaluation model run, based on observations from the population model (e.g., Punt 1993). Model-based management procedures are thus more complicated to evaluate and apply than model-free management procedures, which use the observations but do not require regular stock assessments.

Some lessons learned from work done so far (especially from Kirkwood 1997, De Oliveira et al. 1998, Butterworth & Punt 1999 and McAllister et al. 1999):

- most sets of management objectives include maximising catch, maximising stability of catch and minimising risk,
- model-based procedures perform better (but not spectacularly better) than model-free procedures,
- simple stock assessment models perform better than complex models in evaluations,
- management procedures should be case-specific and not generic,
- management procedures should be evaluated against each other, not against absolute criteria,
- management procedures must be evaluated with respect to specific goals for the fishery,
- it is more important to find a robust management procedure than an optimum one,
- robust means robust to mis-specification of population dynamics and environmental stability, and
- using multiple operating models is valuable.

Cooke (1999) suggested that management procedures must be tested against a wide range of mis-specifications of, or uncertainties about, the underlying reality. Specifically, he suggested testing with a range of productivities, different starting conditions, misreported catches, regime shifts, incorrect stock structure, trends in bias of the abundance indices, alternative stock-recruitment hypotheses, linear or cyclic trends in productivity, and episodic events. He discussed the kind of model that is appropriate for a population model in this process, and how it differs from what is desirable in a stock assessment model.

Patterson et al. (2001) suggested that "In cases where important model uncertainties exist, management procedure simulations can be used to help identify decision rules that are robust to such uncertainty."

2.2 Harvest control rule types

Some of the various harvest control rules we encountered in the literature are described in Section 3.3. All except the constant catch rule use feedback from the population in the form of an abundance index, which in this study is observed CPUE. Rules use this signal in one or more of four ways:

- observed CPUE can be compared with a target level,
- the rate of change in observed CPUE (the gradient) can be calculated over two or more years,
- observed CPUE can be used with an assumed q to estimate biomass for the rule calculation, and
- from two or more years of catch and estimated biomass, current surplus production can be estimated.

Thus there are four main types of harvest control rule based on observed CPUE. Polacheck et al. (1999) suggested that rules involving a constant catch rate applied to estimated biomass probably work the best.

In conjunction with these four types, rules may buffer the target catch (TAC or TACC) resulting from the rule. Buffering occurs in a variety of ways, also seen in the rules described in Section 3.3. Common approaches are:

- combining more than one rule type to form a hybrid rule,
- specifying a sensitivity parameter that modifies the rule's result,
- combining the previous TAC with the rule's resulting target catch,
- limiting the amount of increase and decrease to the TAC permitted in any one change,
- specifying minimum or maximum TACs,
- using thresholds to prevent small changes to TACs,
- using different calculations based on the level of observed CPUE,
- using moving averages to smooth out noise in the CPUE observations, and
- restricting change to alternate years.

Thus, although there are four main types of rule, there would be a very large number of possible combinations of these types with the different buffering mechanisms.

2.3 *Terminology*

We propose that the term "decision rule" be replaced by "management procedure" where the definition of McAllister et al. (1999) has been met (Section 2.1). The term "decision rule" should be used only for procedures that have not been tested through simulation or that do not specify what is to be done upon triggering.

Thus the current NSN and NSC rules would be decision rules; the 1997 and 2002 NSS rules are management procedures.

Below, we use the term "harvest control rule" to indicate the component of a management procedure that we are most interested in testing in this study. The population and other sub-models can be viewed simply as tools used in testing the harvest control rules.

3. MANAGEMENT PROCEDURE DESIGN

An early decision is that between a model-based *vs* a model-free procedure. In the former, the *harvest control rule* depends on the outcome of a formal stock assessment; in the latter no assessment is used explicitly, although target levels and harvest rates may be model-based. Model-based management procedures are much more difficult to evaluate, because the evaluation procedure must mimic the periodic stock assessment using simulated data. Model-free management procedures are generally easier to explain to stakeholders. Stakeholders indicated that they would prefer model-free management procedures, as the current rock lobster decision rules are, rather than model-based ones.

The system of models described below was established on an Excel™ spreadsheet, which has the advantage that a number of different harvest control rules could be observed directly under the same stochastic error terms. Sets of 1000 model runs were easy to make and summarise using macros. Development of C++ version of the model system was started but found not to be necessary.

The model system is run for 115 years – the first 15 years are an adjustment period to allow the population to adjust from its initial position, determined randomly, and fishery indicators are measured over the 100-year period.

3.1 Population model

For looking at rules visually and for making many small sets of evaluations from a large number of rules quickly, we think a simple population simulator is valuable. Here we use the Pella-Tomlinson surplus production model with four parameters:

- *K*, the carrying capacity,
- *r*, the productivity parameter,
- *z*, the shape parameter, and
- *q*, the catchability coefficient (relation between CPUE and vulnerable biomass, or between effort and fishing mortality).

To start the model requires an assumed initial biomass, B_{t0}, and an initial catch, C_{t0}. The initial biomass is drawn from a uniform distribution with specified lower and upper bounds based on tuning (see section 5.5), and initial catch is set to a constant fraction of the biomass.

Given parameter values and an initial biomass, surplus production in the first year *t0* is calculated as:

(1) $$P_{t_0} = Pdev_{t_0}\left(\frac{r}{z}\right)B_{t_0}\left(1-\left(\frac{B_{t_0}}{K}\right)^{z}\right)$$

where

(2) $$Pdev_{t_0} = \exp\left(\varepsilon_{t_0} - 0.5\sigma_{\varepsilon}^{2}\right),$$

and where ε_{t_0} is a random number from $N(0,\sigma_\varepsilon)$.

Biomass in year *t* is calculated by subtracting the catch in year *t*-1 from the sum of production and the biomass in year *t*-1:

(3) $$B_t = B_{t-1} - C_{t-1} + P_{t-1} \qquad t \geq t_0 + 1$$

Catch is determined outside the population model. When the biomass hits an arbitrarily specified low value, this is considered a population crash and is scored as such.

Production in year t is calculated as:

$$(4) \qquad P_t = \left(\frac{r}{z}\right) B_t \left(1 - \left(\frac{B_t}{K}\right)^z\right) e^{Pdev_t - 0.5\sigma_\varepsilon^2}$$

where

$$(5) \qquad Pdev_t = \sqrt{\rho^P}\, Pdev_{t-1} + \sqrt{1-\rho^P}\, \varepsilon_t$$

where ρ^P is a parameter that determines the level of autocorrelation in the process error deviations; this is fixed at an arbitrary value.

3.2 Observation model

The observed CPUE in year t, I_t, is calculated from the given q and B_t with a random observation error, δ_t:

$$(6) \qquad I_t = qB_t \exp\left(\delta_t - 0.5\sigma_\delta^2\right) \quad t \ge t_0,$$

where δ_t is a random number from $N(0,\sigma_\delta)$.

3.3 Harvest control rule models

We found a wide variety of harvest control rule types in the literature. We initially examined the behaviour of a wide selection of these, described below in rough increasing order of conceptual sophistication.

The first rule, a "benchmark", was designed to act as a standard against which the other rules could be compared, and is operated without observation errors, implementation error or implementation lags. Our original plan was to quantify the behaviour of other rules against that of the benchmark, but this implies that the benchmark rule is the best possible rule, which is not true for all indicators.

For all harvest control rules tested except for the benchmark, we incorporated the capability for having a "latent year", where catch cannot be changed in two consecutive years. This is incorporated in the 2002 NSS management procedure to provide more catch stability than would otherwise be the case, and to dampen the action of the rule.

We also incorporated an implementation lag. In current New Zealand rock lobster management, data from fishing year t are analysed, and any resulting management action is discussed, in year t+1 for determining a TAC or TACC for year t+2. We mimicked this situation for all rules except the benchmark.

Starr et al. (1997) and Bentley et al. (unpublished results) experimented with different values of a parameter N, the number of years over which the population indicator – observed CPUE – was averaged. The authors found little effect of this parameter on results. In this study the mean observed CPUE for year t, $\bar{I}_t$, is the moving average of observed CPUE over three years:

(7) $$\bar{I}_t = \frac{1}{3}\sum_{d=t-2}^{t} I_d$$

In each run for a harvest control rule with a final target CPUE, if the initial observed CPUE is less than the final target CPUE, the target CPUE for each of the first 15 years is made to be linear between the initial CPUE and the final target CPUE.

Some rules use a constant harvest rate, called F_n. In this notation, n determines the value of F relative to the shape of a plot of yield *vs* F. At the origin of the plot, yield increases steeply with increasing F, then the slope decreases with increasing F and may become zero (at F_{max}) and then decrease. $F_{0.1}$ is the value of F at the point where the slope is 0.1 times that at the origin. $F_{0.2}$ would be more conservative than $F_{0.1}$ (lower value of F).

Target fishing rate rules require estimates of biomass. For such rules, the estimate was obtained from CPUE divided by q, where q was estimated imperfectly from its true value. For each 115-year run indexed by j, the estimate of q used by the harvest control rule was modified from the true value by simulated observation error:

(8) $$\hat{q}_j = q\exp\left(\varphi_j - 0.5\sigma_\varphi^2\right)$$

where φ_j is from N(0, σ_φ).

Rules differ in how soon they are able to function. For instance, the constant-catch rule can generate a TAC for its first year of the population simulation, but the hybrid rule of Bentley et al. cannot supply a TAC until the sixth year, because it must average the gradient indicator over three years and because of the lag. There was a danger that these differences could bias simulated performance of those rules that take longer to start – for instance, when initial biomass was very low, a fixed initial catch could crash the population before the rule had a chance to operate.

To prevent this, the TAC for the first few years where the harvest rule could not operate was determined by $F_{0.1}$ times B_t. Catch in the year before the harvest control rule first operated was made equal to catch in the year before that, so that the harvest control rule was not precluded by a latent year.

3.3.1 Benchmark rule

The benchmark harvest control rule was designed to provide an indication of how good, theoretically, a harvest control rule could be. We used an F_n rule, and experimented with n to find an acceptable balance between mean catch and other indicators. For this rule only, the observation model supplied perfect information (no observation error on CPUE), perfect knowledge of q was assumed, there was no latent year and no time lag, and the TAC was implemented without error. Autocorrelated process error in the population model was left at the level used for other runs (Table 1 in Section 3.5).

The harvest control rule simply applies the F_n harvest rate to actual biomass:

(9) $$TAC_{t+1} = B_t \frac{F_n}{(F+M)}[1-\exp(-F_n - M)]$$

where M is the instantaneous rate of natural mortality.

3.3.2 Constant catch rule

This rule is included just for comparison with other rules, not as a serious candidate. The harvest control rule is

(10) $$TAC_t = Y$$

where Y is a constant for any variant of the rule.

3.3.3 Rule of Cooke (1999) for baleen whales

The management procedure adopted by the IWC is model-based, but the harvest control rule can be implemented in a model-free procedure. In Cooke's version, the harvest control rule is:

(11) $$TAC_{t+1} = brB_t(B_t/K - a)$$

where a and b are parameters of the rule, r and K are estimated parameters of the surplus-production model. With specified r and K, this rule describes TAC increasing quadratically with biomass, with x-intercept a and slope related to b. We used a linear, model-free version:

(12) $$TAC_{t+2} = \begin{cases} a + b\bar{I}_t/\hat{q}_j & \bar{I} > -\hat{q}_j a/b \\ 0.01 & otherwise \end{cases}$$

where a (<0) is the y-intercept and b (>0) is the slope. Note the implementation lag described in Section 3.3.

3.3.4 Rule of Geromont et al. (1999) for Namibian hake

The procedure described by Geromont et al. (1999) for Namibian hake uses only the observed gradient in a biomass index:

(13) $$TAC_{t+2} = TAC_{t+1}(1 + \lambda s_t)$$

where λ is a rule parameter and s_t is "a measure of the average trend in the abundance index" over four years in their example. The authors didn't specify how s_t is calculated. Two options are: calculate the gradient from the running means of CPUE, or estimate the gradients annually and average those. We tried both approaches and concluded that the first option performed far better than the second, which had population crashes when the first method did not, using the same random numbers. s_t is calculated as:

(14) $$s_t = \frac{\bar{I}_t - \bar{I}_{t-1}}{\bar{I}_{t-1}}$$

3.3.5 Rule of Fournier & Warburton (1988)

This harvest control rule was used in a model-based system, but can be adapted to a model-free system. It uses a target biomass, here translated to target CPUE as explained in section 5.5.

The rule responds to the difference between observed and target CPUE, but buffers the response by restricting the amount of change that can happen in one year. If I_t^* is the target CPUE for year *t*, the rule is:

(15) $$TAC_{t+2} = \min\left\{\max\left\{Max_{decrease}TAC_{t+1}, TAC_{t+1} + s\frac{\left(\bar{I}_t - I_t^*\right)}{\hat{q}_j}\right\}, Max_{increase}TAC_{t+1}\right\}$$

where $Max_{decrease}$ determines how much reduction in TAC is permitted, $Max_{increase}$ determines how much increase will be permitted, and *s* is a sensitivity parameter.

3.3.6 Simple gradient-based rule

We developed a simple rule with no parameters, based on the change in CPUE alone:

(16) $$TAC_{t+2} = TAC_{t+1}s_t$$

where

(17) $$s_t = \bar{I}_t / \bar{I}_{t-1}$$

3.3.7 Gradient rule of Bentley

This rule, developed for this study, has a target CPUE, I^*, and its parameters are a sensitivity parameter, *s*, and a threshold level for change, *T*. The rule calculates two intermediate values, *R* and *D*.

(18) $$R_t = \bar{I}_t / \bar{I}_{t-x}$$

where *x* is 1 if the TAC changed in year *t*-1, 2 if the TAC was changed in the year before that, and so on.

(19) $$D_t = 1 + (\bar{I}_t / I_t^* - 1)s$$

The amount by which the TAC changes, Z_t, is:

(20) $$Z_t = \begin{cases} 1 & abs(1 - R_t D_t) < T \\ R_t D_t & abs(1 - R_t D_t) \geq T \end{cases}$$

then

(21) $$TAC_{t+2} = Z_t TAC_{t+1}$$

3.3.8 Rule of Bentley et al. for rock lobsters

This harvest control rule is a hybrid between a target-based rule such as that of Fournier & Warburton (1988) and a gradient-based rule such as that of Geromont et al. (1999). The rule uses target CPUE, I_t^*, and parameters are *S*, the scale factor, and *W*, the relative weight given to the status indicator *vs* the gradient indicator. Other variables used are A_t^s, the status indicator that measures the displacement of the observed CPUE from the target CPUE in year *t*; and A_t^g, the gradient indicator that measures the gradient of CPUE in year *t*.

We calculate the status indicator, A_t^s and the gradient indicator, A_t^g as:

(22) $$A_t^s = {I_t}/{I_t^*} - 1$$

(23) $$A_t^g = \left(\left(I_t - I_{t-1}\right)/I_{t-1}\right) - \left(\left(I_t^* - I_{t-1}^*\right)/I_{t-1}^*\right)$$

The moving averages of A_t^s and A_t^g over 3 years are calculated as:

(24) $$\overline{A}_t^s = \frac{1}{3}\sum_{d=t-2}^{t} A_d^s$$

and

(25) $$\overline{A}_t^g = \frac{1}{3}\sum_{d=t-2}^{t} A_d^g$$

The moving averages of gradient indicator and the status indicator are combined using with the relative weight, *W*, to give the combined indicator:

(26) $$A_t = W\overline{A}_t^s + (1-W)\overline{A}_t^g$$

The response, R_t, is calculated from the combined indicator with the scale factor:

(27) $$R_t = SA_t$$

The catch multiplier, Z_t is defined as:

(28) $$Z_t = \begin{cases} 1 & abs(R_t) < 0.05 \\ 1 + R_t & 0.05 \le abs(R_t) < 0.25 \\ 1.25 & R_t \ge 0.25 \\ 0.75 & R_t \le -0.25 \end{cases}$$

The TAC from the rule is:

(29) $$TAC_{t+2} = Z_t TAC_{t+1}.$$

3.3.9 Rule of Haist (2002) for hoki

This rule is based on estimated surplus production and is buffered by combining the result of the rule with the previous TAC. The rule uses a target CPUE, I_t^*, and parameters are a scale factor, s, the relative weight given to the previous TAC, Δ, and a minimum CPUE, $I_{\min}$. The proportion f_t, of the estimated surplus production that is allocated to the catch is calculated as:

$$(30) \qquad f_t = \begin{cases} 0 & \bar{I}_t \le I_{\min} \text{ or } \hat{P}_t < 0 \\ \dfrac{\bar{I}_t - I_{\min}}{I_t^* - I_{\min}} & I_{\min} < \bar{I}_t < I_t^* \\ 1 + s\left(\dfrac{\bar{I}_t}{I^*}\right) & \bar{I}_t > I_t^* \end{cases}$$

where $\hat{P}_t$ is the estimated surplus production in year t. Surplus production is estimated from a rearrangement of (3):

$$(31) \qquad P_t = B_{t+1} - B_t + C_t$$

To do this, biomass in each year is estimated from CPUE and an estimated q, observed with error as described above:

$$(32) \qquad \hat{P}_t = \left(\bar{I}_{t+1} - \bar{I}_t\right)/\hat{q}_j + C_t$$

Then the TAC for year $t+2$ is defined as:

$$(33) \qquad TAC_{t+2} = (\Delta)TAC_{t+1} + (1-\Delta)f_t\hat{P}_t.$$

3.3.10 Rule of Geromont & Glazer (1998) for South African hake

This is an F_n rule buffered by combining the estimated F_n catch with the previous TAC:

$$(34) \qquad TAC_{t+2} = \Delta TAC_{t+1} + (1-\Delta)C_t^{F_n},$$

where $C_t^{F_n}$ is the estimated F_n catch for year t. This is determined by estimating biomass from CPUE and estimated q:

$$(35) \qquad C_t^{F_n} = \left(\bar{I}_t/\hat{q}_j\right)\frac{F_n}{(F_n + M)}\left[1 - \exp(-F_n - M)\right].$$

3.3.11 Bentley's F_n rule

This rule, developed for this study, uses an F_n catch based on biomass projected two years ahead from the trend in CPUE. The rule is buffered by previous TAC using a Δ parameter. Future biomass is predicted from current CPUE modified by two years of the current gradient in CPUE:

(36) $$\hat{B}_{t+2} = \left[\bar{I}_t + 2(\bar{I}_t - \bar{I}_{t-1})\right]/\hat{q}_j$$

which simplifies to

(37) $$\hat{B}_{t+2} = \left(3\bar{I}_t - 2\bar{I}_{t-1}\right)/\hat{q}_j$$

The F_n catch, $C^{F_n}_{t+2}$, is calculated from F_n and $\hat{B}_{t+2}$ as in (9), and the TAC given by:

(38) $$TAC_{t+2} = \Delta TAC_{t+1} + (1-\Delta)C^{F_n}_{t+2}.$$

3.3.12 Rule of Punt & Smith (1999) for gemfish

These authors used the rule in a model-based system, but it can also be used in a model-free system. It is an F_n strategy buffered with minimum and maximum TACs, TAC_{min}, TAC_{max}, and by separate limits on the amount by which TAC can change in one year, $Max_{decrease}$, and $Max_{increase}$. The maximum and minimum TACs permitted by the rule for year t+2 are:

(39) $$Q^{max}_{t+2} = \min\left\{TAC_{max}, Max_{increase}TAC_{t+1}\right\}$$

(40) $$Q^{min}_{t+2} = \max\left\{TAC_{min}, Max_{decrease}TAC_{t+1}\right\}$$

These are the bounds for the TAC. The TAC is set to $C^{F_n}_t$ (35), if it is between the bounds, or to a bound if outside the bound:

(41) $$TAC_{t+2} = \begin{cases} Q^{max}{}_{t+2} & Q^{max}{}_{t+2} < C^{F_n}_t \\ C^{F_n}_t & Q^{min}{}_{t+2} \le C^{F_n}_t \le Q^{max}{}_{t+2} \\ Q^{min}{}_{t+2} & C^{F_n}_t < Q^{min}{}_{t+2} \end{cases}$$

3.3.13 Rule of De Oliveira et al. (1998) for sardines

This rule is another buffered F_n catch rule, with parameters n, TAC_{min}, TAC_{max}, and $Max_{decrease}$. The rule operates as for the rule of Punt & Smith (1999), but the maximum allowable TAC, Q^{max}, is simply TAC_{max}. The rule then acts as in (41).

3.3.14 Rule of Baldursson et al. (1996) for Icelandic cod

This rule is an arbitrary function of estimated biomass. The TAC is set as:

(42) $$TAC_{t+2} = \min\left(\max\left(Q_{\min,t+2}, \max\left(Q_{\max,t+2}, a\left(\bar{I}_t/\hat{q}_j\right) - b\right)\right), C^{F_{high}}_t\right)$$

where

(43) $$Q_{\min,t+2} = \max(TAC_{\min}, (1-D)TAC_{t+1})$$

and

$$(44) \qquad Q_{\max,t+2} = \min(TAC_{\max}, (1+D)TAC_{t+1})$$

In this rule, the F_{high} catch is intended as an upper limit, associated with a high value of F. With parameters set at 300 and 1000 for minimum and maximum TAC, $D = 0.1$, $a = 0.5$ and b=2000, and $F_{high} = 0.6$, the equilibrium catch *vs* biomass relation is shown in Figure 2.

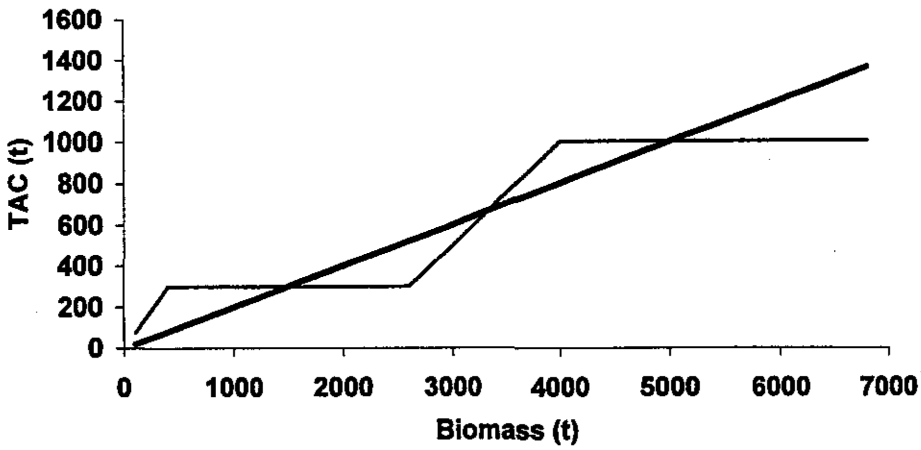

Figure 2: Equilibrium TAC *vs* biomass from the rule of Baldursson et aL (1996). The bold line shows the catch derived from $F = 0.20$.

This is a counter-intuitive rule: when biomass is very small, the TAC is larger than $C_t^{F_n}$ obtained with a reasonable F (0.20 in Figure 2), it is less than $C_t^{F_n}$ as biomass increases, and flips back and forth from being less than or greater than $C_t^{F_n}$.

3.4 Implementation model

The implementation model translates the TAC specified by the harvest control rule into an actual catch. Two processes are involved: first, catch must be limited when the TAC is large relative to the biomass available, and this is done by specifying a maximum exploitation rate, U_{max}.

Second, the actual catch will seldom be exactly the TAC. Fishers may not take the TAC because of bad weather, low prices, alternative endeavours or whatever. In the actual rock lobster fishery at present, only the legally operating commercial fishery is capped by the TAC (through an explicit TACC, a sub-set of the TAC). Recreational, traditional and illegal fisheries are not capped and so their catches may bear no relation to what is allowed for them in the TAC.

For the base case we model the TAC as a target for all the fisheries combined, and incorporate process error:

$$(45) \qquad C_t = \min\left\{U_{\max} B_t, TAC_t e^{Cdev_t - 0.5\sigma_\vartheta^2}\right\}$$

where $Cdev_t$ is the deviation from the TAC, determined from

(46) $Cdev_t = \vartheta_t$ for $t = t0$, otherwise

(47) $$Cdev_t = \sqrt{\rho^C} Cdev_{t-1} + \sqrt{1-\rho^C} \vartheta_t$$

where ρ^C is a parameter that determines the auto-correlation in catch deviations and ϑ_t are from $N(0, \sigma_\vartheta)$.

In robustness testing, we model unreported catch and uncontrolled catch by changing the implementation model as described in Section 3.7.3.9.

3.5 Base case specifications

This section describes all the variable and parameter choices that we used in base case comparisons of the various harvest control rules. Other changes made for robustness testing will be described in Section 3.7.

Table 1: Variable and parameter choices that were used in base case comparisons of the various harvest control rules.

Description	Equation	Symbol	Value
run length		-	115 yr
Population model			
carrying capacity	(1)	K	6985 t
productivity	(1)	r	0.251
shape	(1)	z	0.02
catchability	(1)	q	2.72E-04
natural mortality rate	(9)	M	0.10
initial biomass	(1)	B_{t0}	U(1300,5200)
initial catch	-	C_{t0}	$0.1623B_{t0}$
sigma for process error	(2), (5)	σ_ε	0.20
autocorrelation in process error	(5)	ρ^P	0.25
population crash	-	-	$B_t < 256$ t
Observation model			
sigma for observation error	(6)	σ_δ	0.10
Harvest control rules			
number of years' CPUE averaged	(7)	N	3
target biomass	-	B^*	3250 t
target CPUE	(15)	I^*	0.884 kg/lift
sigma for estimated catchability	(8)	σ_φ	0.10
Implementation model			
maximum exploitation rate	(45)	U_{max}	0.75
sigma for catch deviations	(45), (47)	σ_ϑ	0.08
autocorrelation in catch deviations	(47)	ρ^C	0.50

The basic population model parameters came from estimation (see Section 4). M has been previously assumed at 0.10 for *Jasus edwardsii* when not estimated in assessments. The "target biomass" was 1.25 times the estimated B_{msy} (Section 4.5), and target CPUE was estimated as q times that.

Initial biomass was drawn from a uniform distribution, centred on the arbitrary "target biomass" of 3250 t, to give some contrast in starting points. The standard deviation and autocorrelation of process error were chosen arbitrarily - the chosen ρ^P gives an average autocorrelation of 0.5. A typical string of process error deviations in arithmetic space is shown in Figure 3.

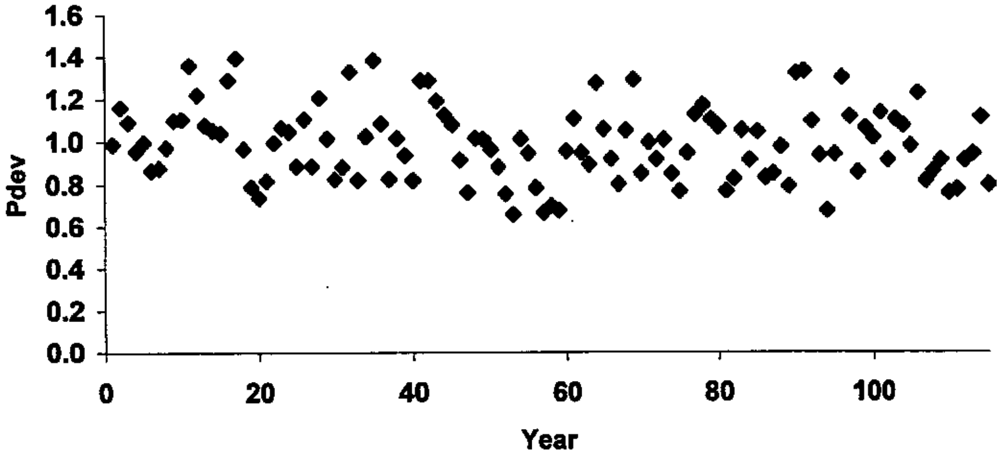

Figure 3: A typical string of process error deviations from a 115-year run.

The standard deviation for observation error in CPUE was chosen arbitrarily, but with consideration of Francis et al. (2001) and our experience in rock lobster assessments. Figure 4 shows the simulated actual and observed CPUE from a typical run.

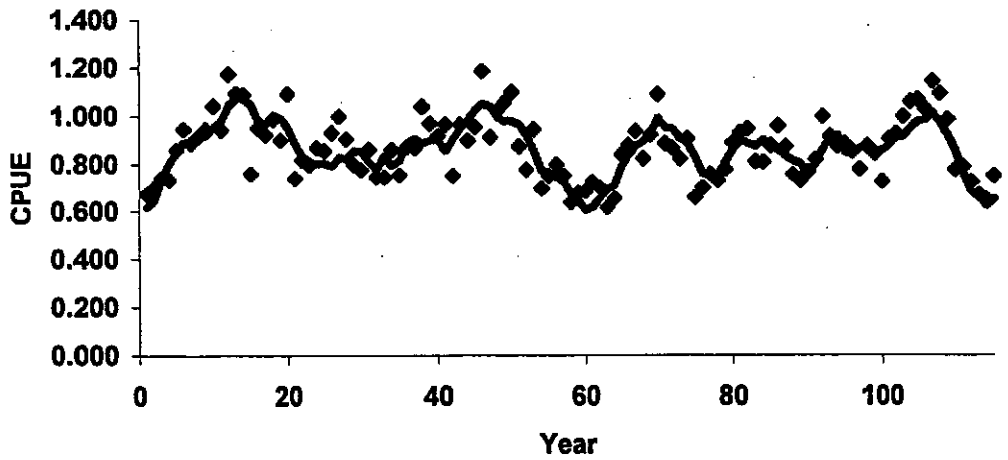

Figure 4: Simulated actual (line) and observed (diamonds) CPUE from the population and observation models respectively, from a typical 115-year run.

The standard deviation of estimation error for q was chosen arbitrarily after inspecting plots of its effect – it gives an average of 8% error in $\hat{q}_j$ and no autocorrelation is generated. Figure 5 shows the generated error around q in a typical run.

Maximum exploitation rate was arbitrary. The standard deviation and autocorrelation parameters for implementation error were chosen arbitrarily: Figure 6 shows the actual catch *vs* TAC in a typical run.

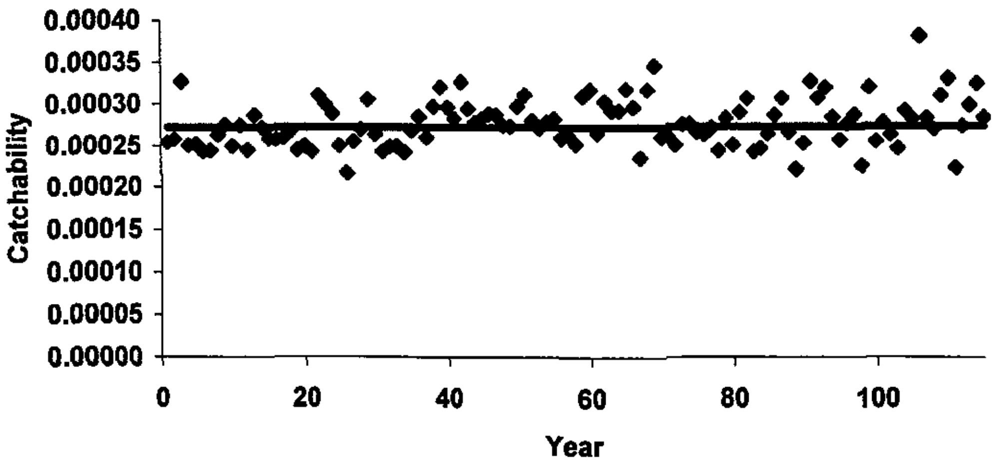

Figure 5: Actual q (line) and observed (diamonds) catchability, $\hat{q}_j$, from a typical 115-year run.

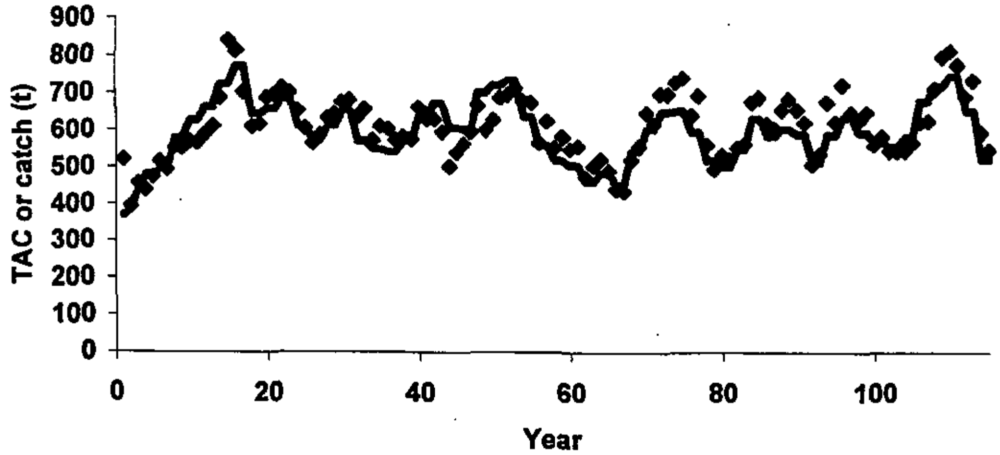

Figure 6: TAC (line) from the harvest control model and observed catch (diamonds) from the implementation model from a typical 115-year run.

3.6 Performance indicators

Performance indicators are measured over the 100-year period after the 15-year stabilisation period. They are:

3.6.1 Yield indicators

- We calculate mean catch from each run ("*Cav*").
- We calculate minimum catch from each run ("*Cmin*").

3.6.2 Risk indicators

- Minimum biomass is calculated from each run ("*Bmin*").

- A low biomass index is calculated ("*lowB*") to reflect the extent to which biomass is below the arbitrary threshold of 3250 t:

(48) $$lowB = \sum_{t=16}^{t=115} \left(\max\{0, 3250 - B_t\}\right)^2 \bigg/ 1000000$$

- The percentage of crashes in a set of runs is calculated "*%Crash*").

3.6.3 Stability indicators

- A gross measure of variation in biomass during the run is the maximum biomass minus the minimum ("*Bdiff*")

- We also calculate the average annual absolute percentage variation in TAC ("*AAV*"):

(49) $$AAV = 100 \sum_{t=16}^{t=115} \frac{|TAC_t - TAC_{t-1}|}{TAC_{t-1}}$$

3.6.4 Abundance indicator

- Mean biomass is calculated from each run ("*Bav*").

3.7 Testing harvest control rules

The various harvest control rules were tested in 100-year stochastic runs on an Excel™ spreadsheet. For all sets of runs, the random numbers used in population, observation and implementation randomness were the same for all control rules, so that differences among rules do not reflect differences in the stochastic environment. All sets of comparisons were made with 1000 runs.

3.7.1 Preliminary rule testing

We first compared all the rules described in Section 3.3. We experimented informally with the various parameters for each rule, to find a specific rule that appeared to perform adequately, but at this stage we did not do serious tuning for any rule.

3.7.2 Testing of tuned rules

After the initial experiments, we eliminated some rules that did not perform adequately. before doing this we did further experimentation with their parameter values to ensure that we had not failed to find a good example of the rule. For the remaining rules, more careful and systematic exploration was made to ensure that good parameter values had been chosen. We then ran a set of runs comparing the tuned rules. All the work described so far in this section used base case assumptions.

3.7.3 Robustness trials

We conducted robustness trials by making substantial changes to the population, observation or implementation models in some way, mimicking possible differences between reality and the assumed model reality. Deviations from the base case were selected arbitrarily. These trials were all conducted with the latent year implemented.

Changes to the population model were:

1. randomly varying parameters of the population model
2. introducing a lag in population productivity,
3. varying productivity in a long cycle,
4. introducing a regime shift in first 50 years of the run,
5. introducing random mortality events, and
6. increasing the stochastic errors.

The change to the observation model was:

7. a systematic increase in q over the first 50 years of a run.

Changes to the implementation model were:

8. causing catches to exceed the TAC, and
9. incorporating a substantial catch outside the scope of the harvest control rule.

These changes were implemented one at a time from the base case, not as sequential changes.

3.7.3.1 Varying population model parameters

The harvest control rules had been tuned to a specific population model. In reality, the population model used to tune management procedures might be incorrect, i.e., values other than the true values might be used based on an imperfect stock assessment. To explore the effect of this, we randomly chose the parameters r, K, q and z from normal distributions at the beginning of each run. The distributions for r, K, and q had means equal to the base case parameters (see Table 1) and standard deviations of 0.05, 200 and 5E-5 respectively. The distribution for z was uniform between 0.001 and 0.500.

3.7.3.2 Productivity lag

In this robustness trial, productivity in year t was related to the biomass in year t-2 rather than in t-1 as in the base case (cf Eq. 4):

(50) $$P_t = Pdev_t\left(\frac{r}{z}\right)B_{t-2}\left(1-\left(\frac{B_{t-2}}{K}\right)^z\right)$$

3.7.3.3 Variable productivity

Productivity was made to vary over time by varying K from the base K (called $\widehat{K}$) with a cosine curve with one cycle over the 100-year run:

(51) $$K_t = \widehat{K}\left(1+0.4\cos\left(0.6+\frac{t}{15.5}\right)\right)$$

The factor 15.5 scales the periodicity of the cosine wave; it was set to give one complete cycle over 100 years when the cosine function takes radians as the argument. The factor 0.6 was chosen to begin the cycle at $\hat{K}$ in year 16. Figure 7 shows the time series of K_t used in this test.

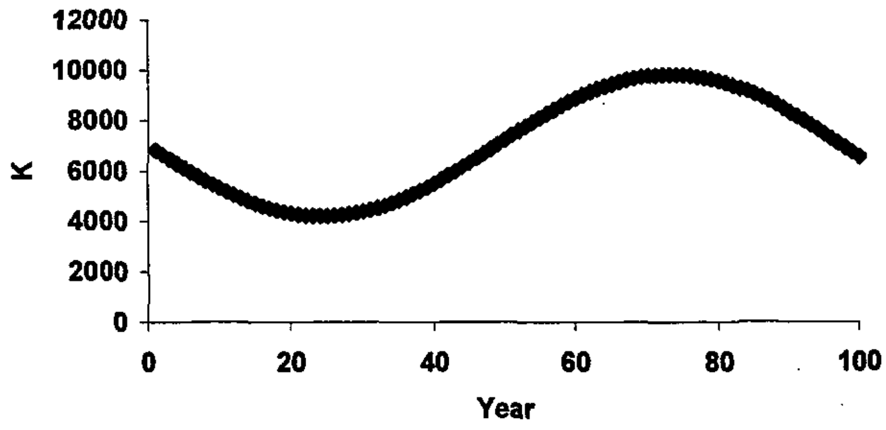

Figure 7: The time series of K_t used in the robustness test in which K varied over time.

3.7.3.4 Regime shifts

We simulated a regime shift by changing the value of K to a new value, K', drawn from a normal distribution with a mean of the base case value, $\hat{K}$, and a c.v. of 0.25. This happened in a year randomly chosen between years 16 and 65 (the first half of the 100 years that are summarised in the fishery indicators).

3.7.3.5 Episodic mortality

In this test, in each year there was a probability of 0.025 that total biomass would experience a 40% mortality. This simulates random episodic events.

3.7.3.6 Increasing stochastic error

In this trial we doubled σ_ε, σ_δ and σ_θ in the population, implementation and observation models. These are the standard deviations for the random productivity process error, CPUE observation error, and catch implementation error respectively. Because this change turned out to have a large effect, we then made these changes singly to see which individual value had caused the change.

3.7.3.7 Increasing *q*

This test simulated the situation where effective effort increased over time because of technological change but the change was unnoticed. The change increases apparent CPUE.

We simulated a 0.5% annual increased in the actual value of q. For the benchmark rule we wished to compare the effect of this change – in the base case the benchmark rule uses perfect information, but in this trial we drove the benchmark rule with the original estimate of q.

3.7.3.8 Non-reported catches

In this trial the implementation model was changed to cause the actual catch, unless constrained by the maximum exploitation rate, to be larger than the TAC by a random factor drawn for each run from a normal distribution with a mean of 1.10 and a standard deviation of 0.075.

3.7.3.9 Additional uncontrolled catches

In this trial the implementation model generated two catch series: commercial catch (controlled by a limit called the TACC) and non-commercial catch (uncontrolled). Total catch was limited when the TAC (here the sum of commercial and non-commercial catches) was large relative to the biomass available, by specifying the maximum exploitation rate, U_{max}.

The actual total catch was calculated as:

(52) $$C_t^{actual} = \min\{U_{\max} B_t, C_t^1 + C_t^2\},$$

where C_t^1 and C_t^2 are the "targets" for commercial and non-commercial catches respectively. The target for commercial catch is related to the TACC, as in the base case implementation model (section 5.4):

(53) $$C_t^1 = TACC_t e^{Cdev_t^1 - 0.5\sigma_{\vartheta^1}^2}$$

We model the recreational catch as a function of biomass and allowed for the possibility of systematic increase:

(54) $$C_t^2 = e^{Cdev_t^2 - 0.5\sigma_{\vartheta^2}^2} (\alpha + \beta B_t)(1 + \nu)^t,$$

where α and β are constants determining the range of non-commercial catch, ν is the annual rate of increase in non-commercial catch, and $Cdev_t^1$ and $Cdev_t^2$ are annual stochastic deviations from the commercial and non-commercial catch, determined from

(55) $$Cdev_t^i = \vartheta_t^i \qquad \text{for } t = t0, \text{ otherwise}$$

(56) $$Cdev_t^i = \sqrt{\rho^i}\, Cdev_{t-1}^i + \sqrt{1 - \rho^i}\, \vartheta_t^i$$

where ρ^i is a parameter that determines the auto-correlation in catch deviations, ϑ_t^i are from $N(0, \sigma_{\vartheta^i})$ and i is 1 for commercial catch and 2 for non-commercial catch.

The actual commercial and non-commercial catches are taken as C_t^1 and C_t^2 when their sum is less than $U_{\max} B_t$, otherwise the catch is calculated from the proportions of C_t^1 and C_t^2:

(57) $$AC_t^i = U_{\max} B_t \left(\frac{C_t^i}{C_t^1 + C_t^2} \right)$$

where i is 1 for commercial catch and 2 for non-commercial catch.

We used four scenarios: all with $\beta = 0.07$ and with α=75 t or 200 t, and with the annual rate of increase of the non-commercial catch function 0% or 1%. The values for the intercept gave initial non-commercial catches of 285 and 410 t respectively when biomass was 3000 t. In all these runs, the F_n rules used a value of $F = 0.1022$, determined from a rough balance at the estimated current values for the NSN stock.

4. DETERMINING BASE CASE POPULATION MODEL PARAMETERS

To determine appropriate values for the parameters K, r, z and q for use in the population dynamics model, we fitted an ADModel Builder version™ to the NSN data (i.e., CRA 1 and CRA 2) and to the estimated exploitation rates from the most recent assessment (Breen et al. 2001).

4.1 Data

We used the CPUE, exploitation rate and the catch series from the 1999 stock assessment (Breen et al. 2001). These data were re-organised into annual values from the 6-month period data used by the assessment. Data were summarised as follows: the catch series was annualised by summing the two period catches; the CPUE series was annualised by dividing the annual catch by the annual effort. Annual effort was found by summing the period effort, in turn calculated by dividing catch by CPUE for each period. The exploitation rate was calculated by dividing the annual catch by the model's estimated recruited biomass at the beginning of the autumn-winter period.

Parameters estimated were K, r, z and q. We also estimated production deviations ε_{t_0}. We fixed the parameter ρ^P to an arbitrary low value, 0.0001 (i.e., no correlation of production deviations was assumed). All parameters except for the ε_{t}, which had a normal prior with a mean of zero and σ_ε of 0.20, were assumed to have uniform priors with wide bounds.

4.2 Initialisation

The model initialises biomass at the beginning of estimation. Biomass in the first year is calculated from the given data values for the catch and the exploitation rate.

$$(58) \qquad B_{t_0} = \frac{C_{t_0}}{U_{t_0}}$$

Then the surplus production in the first year is calculated as in (1) and (2).

4.3 Dynamics

From the second year, the biomass is calculated as in (3), and the production is calculated as in (4) and (5).

From the calculated biomass and the given catch, we can calculate the predicted exploitation rate, $\hat{U}_t$, and the predicted CPUE, $\hat{I}_t$. The predicted exploitation rate is:

(59) $$\hat{U}_t = \frac{C_t}{B_t}$$

and the predicted CPUE is:

(60) $$\hat{I}_t = B_t\hat{q}\text{, where}$$

(61) $$\hat{q} = \exp\left(\frac{1}{m}\sum_t \ln\left(\frac{I_t}{B_t}\right)\right),$$

where I_t is observed CPUE and B_t is model biomass in year t respectively and m is the number of years of observed CPUE data.

4.4 Likelihood

Normal likelihood was used to calculate the fits to the CPUE and exploitation rate. The negative log-likelihood for the CPUE is:

(62) $$L = \sum_t \left(\frac{\left(\ln(\hat{I}_t) - \ln(I_t)\right)^2}{2\sigma_I^2} + \ln(\sigma_I) + 0.5\ln(2\pi) \right),$$

where σ_I is the standard deviation of the CPUE error distribution, which was assumed to be 0.02. The likelihood for the U_t was analogous to (62), and σ_U was assumed to be 0.10. The values for σ_I and σ_U were chosen so that the standard deviation of the standard residuals for each data set was approximately 1.

The negative log-likelihood of the ε_t was given by a normal likelihood,

(63) $$L = \frac{\sum_t (\varepsilon_t)^2}{2\sigma_\varepsilon^2} + \ln(\sigma_\varepsilon) + 0.5\ln(2\pi)$$

4.5 Estimated parameters

Table 2 shows the resulting parameter value estimates (as modes of the joint posterior distributions, MPDs) and the function component values from the various data sets; Figures 8 and 9 show the fits to the CPUE and exploitation rate respectively. The model obtains a very tight fit to the CPUE and a reasonable fit to the assessment model's estimated exploitation rates.

From these values, the deterministic value of *MSY* was 639 t, B_{MSY} was 2600 t, and the CPUE associated with B_{MSY} was 0.7072 kg/pot-lift.

Table 2: Estimated parameter values and likelihoods. "SDSDRs" are the standard deviations of standardised residuals.

Parameter	Value
r	0.251
K	6985
z	0.020
$\hat{q}$	0.000272
Likelihood	
CPUE	-42.35
Exploitation rate	-36.62
Production deviations	53.21
Total	-25.76
SDSDRs	
CPUE	1.32
Exploitation rate	1.12

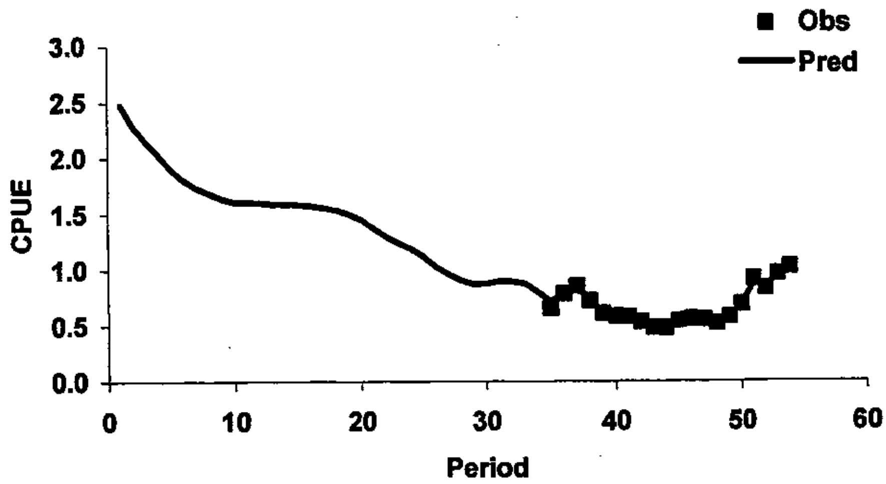

Figure 8: Fit of the Pella-Tomlinson model (line) to the observed NSN CPUE (squareds).

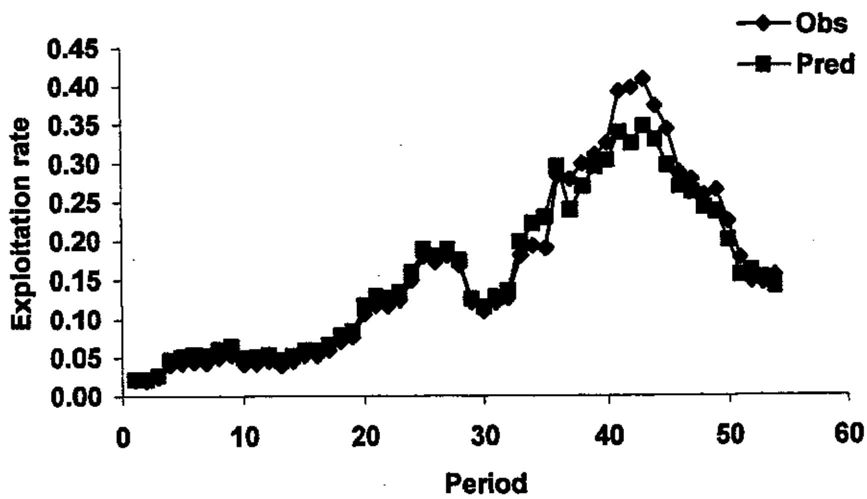

Figure 9: Fit of the Pella-Tomlinson model (squares) to the estimated exploitation rate series from Breen et al. (2001) (diamonds).

4.6 Reference fishing mortality rates $F_{0.1}$ and F_n

Some harvest control rules use reference *F* values. We obtained these from a simple yield-per-recruit analysis. Values for growth parameters were taken from the NSN stock assessment in 1999 (Breen et al. 2001), and are shown in Table 3.

Table 3: Parameters used for the yield-per-recruit calculation.

Parameter	Description	Value
M	natural mortality rate	0.10
male growth		
Ginf	asymptotic tail width	83.2
GmaxMoult	size at which moulting becomes annual	162.4
Gslope	slope of increment *vs* initial size	-0.094
MLS	minimum legal size	54
a	intercept for weight *vs* tail width	0.00416
b	slope for weight *vs* tail width	2.935
female growth		
Ginf	asymptotic tail width	72.6
GmaxMoult	size at which moulting becomes annual	57.4
Gslope	slope of increment *vs* initial size	-0.178
MLS	minimum legal size	60
a	intercept for weight *vs* tail width	0.0130
b	slope for weight *vs* tail width	2.545

Annual growth was calculated for each sex separately as described by Breen et al. (2001). The NSN stock assessment reported semi-annual growth, so we calculated two periods' growth increments and added them to get annual increment. In the NSN assessment model, each year has a primary growth period and a secondary growth period. In the primary growth period, every rock lobster grows; in the secondary growth period, only animals less than an estimated size (*GMaxMoult*) grow. A large value of *GMaxMoult* for males (162.4 mm) means they all moult twice a year; only small females moult twice per year.

The yield-per-recruit model is age structured, with a population assumed to be in equilibrium with recruitment and natural and fishing mortality. Recruited biomass is calculated as:

(64) $$B = \sum_{G_a \geq MLS} N_a w_a ,$$

where G_a is the tail width at age *a* and N_a is the number of animals at age *a*, which is an arbitrary *R0* at age 1 and

(65) $$N_a = N_{a-1}\exp(-M) \qquad \text{for } G_{a-1} < MLS,$$
$$N_a = N_{a-1}\exp(-(M+F)) \qquad \text{for } G_{a-1} \geq MLS$$

for all other ages. The fishing mortality, *F*, is specified, and for females is 0.5 times male fishing mortality rate to account for the protection given to berried females..

Equilibrium catch (or yield) is calculated as:

(66) $$C = B\left(\frac{F}{F+M}\right)(1-\exp(-(F+M))).$$

Values are shown in Table 4.

Table 4: Reference fishing mortality rates from yield-per-recruit modelling of the NSN stock.

n	F_n
0.05	0.2356
0.06	0.2151
0.07	0.1984
0.08	0.1845
0.09	0.1726
0.10	0.1623
0.15	0.1251

There was some incompatibility between these reference values and those obtained from a plot of yield *vs* equilibrium biomass from the surplus-production model. In the surplus-production version, $F_{0.1}$ was 0.2147. Conversely, the value of r that gave $F_{0.1}$ = 0.1623 for the surplus production model was 0.1919, less than the estimated r = 0.251. This does not really affect the study.

5. RESULTS

5.1 Preliminary explorations

We initially explored the behaviour of each of the rules described in Section 3.3 to see how well they worked after informal tuning of their parameters.

5.1.1 Benchmark rule

Figure 10 shows a typical run from the benchmark rule. TAC lags behind biomass slightly: although there is no implementation lag, the benchmark rule must estimate biomass from observed CPUE from the previous year. This is an F_n rule: experimentation with the value for n suggested that 0.06 gives mean biomass, *Bav*, closest to the target of 3250 t; mean catch, *Cav*, appears to be just above 600 t.

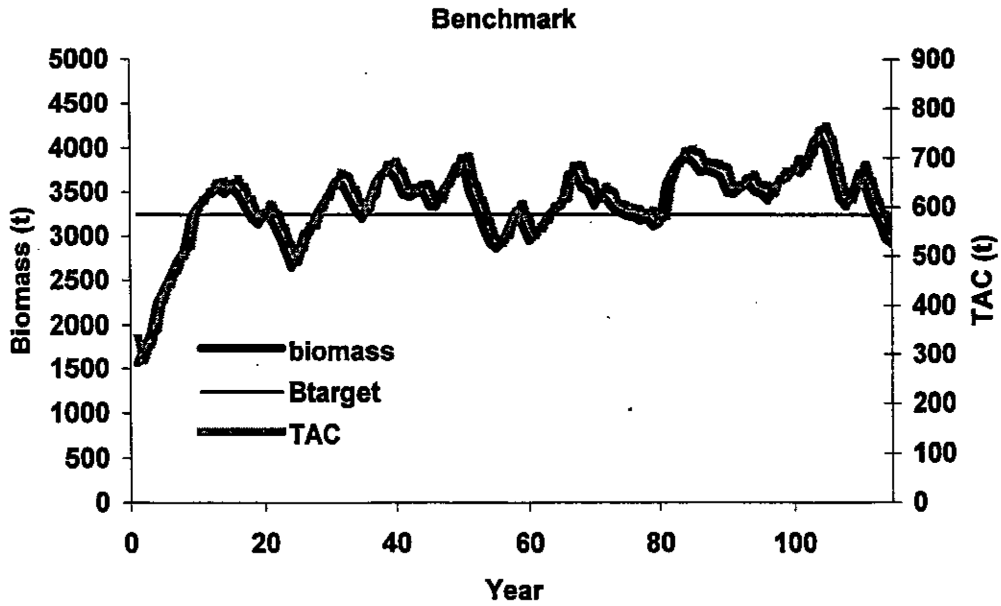

Figure 10: A typical run with the benchmark rule.

5.1.2 Constant catch rule

The constant catch rule behaves differently with different TACs. With a low TAC, *Bav* stabilises near 5000 t. With increasing TAC, *Bav* is lower and the fluctuations in biomass become more pronounced (Figure 11). Crashes begin to appear with low frequency when the TAC is around 500-550 t, and the frequency of crashes increases sharply at higher TACs (Figure 12).

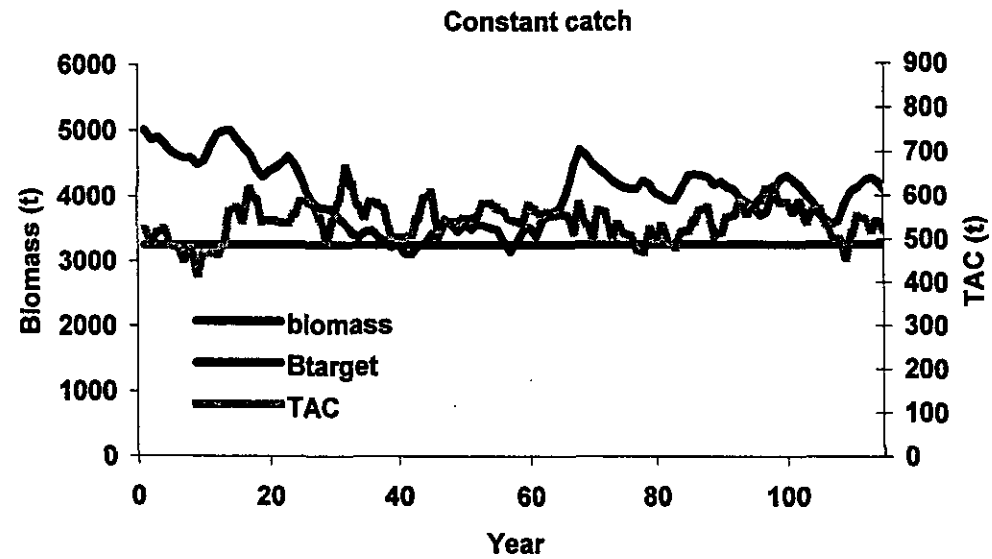

Figure 11: Showing fluctuations in biomass and actual catch from a typical run under the constant catch rule with a TAC of 550 t.

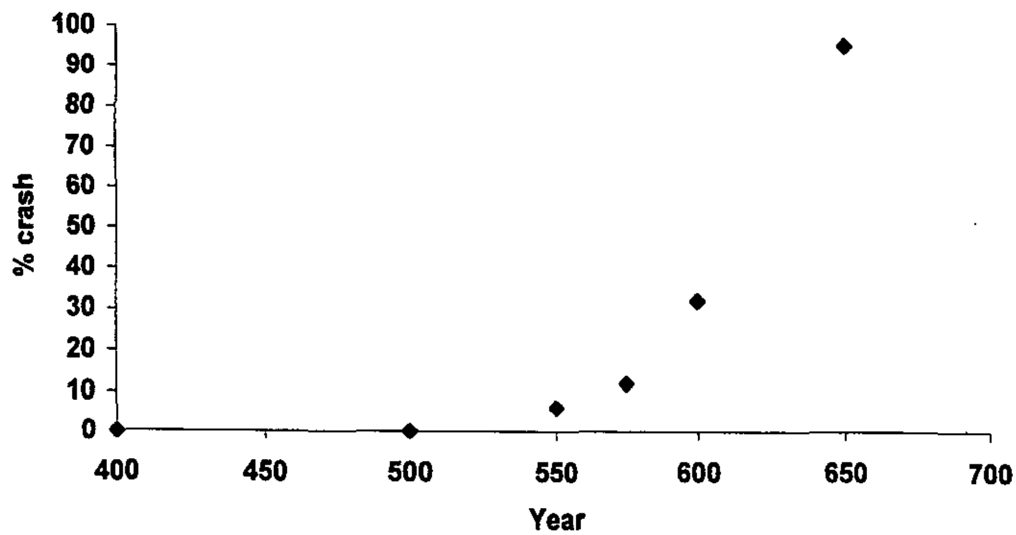

Figure 12: The percentage of crashes in 100 runs using the constant catch rule, plotted as a function of the constant TAC.

5.1.3 Simple gradient-based rule

This rule has no parameters. The rule behaves well without a latent year, although in some cases the scale of fluctuations seems large (Figure 13). Otherwise, the rule seems to deliver mean biomass of roughly the target and a good mean catch, and we saw no crashes in this phase.

However, with a latent year, the rule does not perform well: TAC becomes disconnected from biomass (Figure 14), and biomass crashes about 15% of the time.

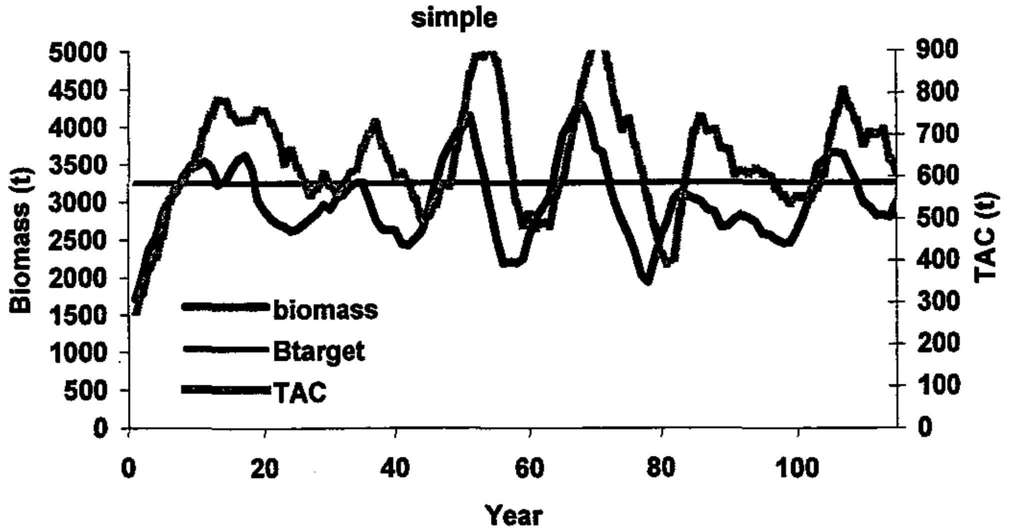

Figure 13: Biomass and TAC (grey) from the simple rule with no latent year, showing an extreme case with respect to fluctuations - most runs were better behaved than this.

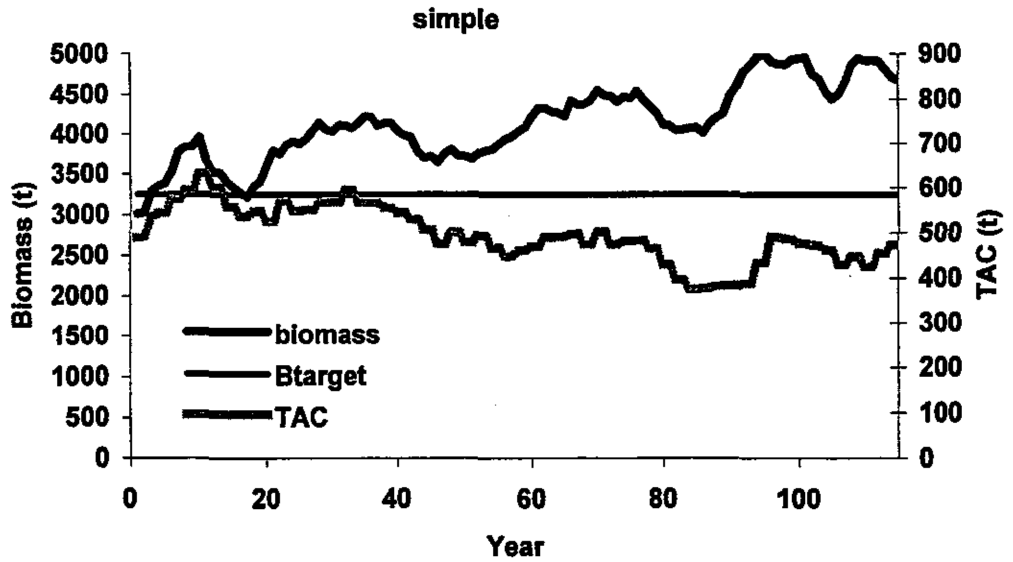

Figure 14: Biomass and TAC (grey) from the simple rule with a latent year, showing the disconnection of TAC from biomass.

5.1.4 Rule of Cooke (1999) for baleen whales

This rule has parameters a and b – the intercept and slope of the plot of TAC against biomass. To obtain *Bav* apparently near the target we chose parameters $a = -200$ and $b = 0.25$. The latent year does not seem to make a substantial difference to the rule. We saw no crashes. Stability of the rule is good in many runs and poor in other runs (Figures 15 and 16). These examples show the effect of the implementation lag: catch is often decreasing when biomass is increasing, or vice versa, and cycles are created that are unrelated to the underlying production deviations.

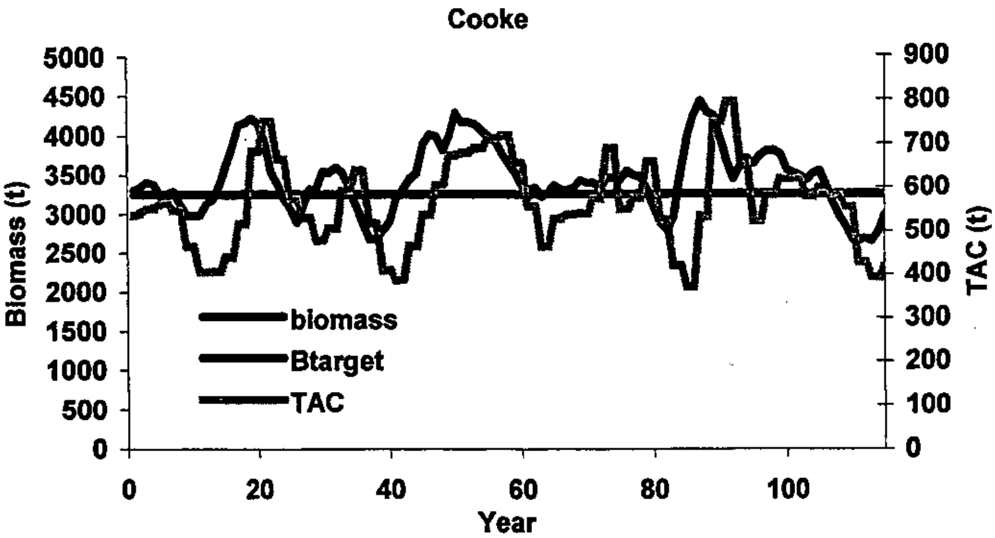

Figure 15: Biomass and TAC (grey) from a good example of the Cooke rule, using a latent year.

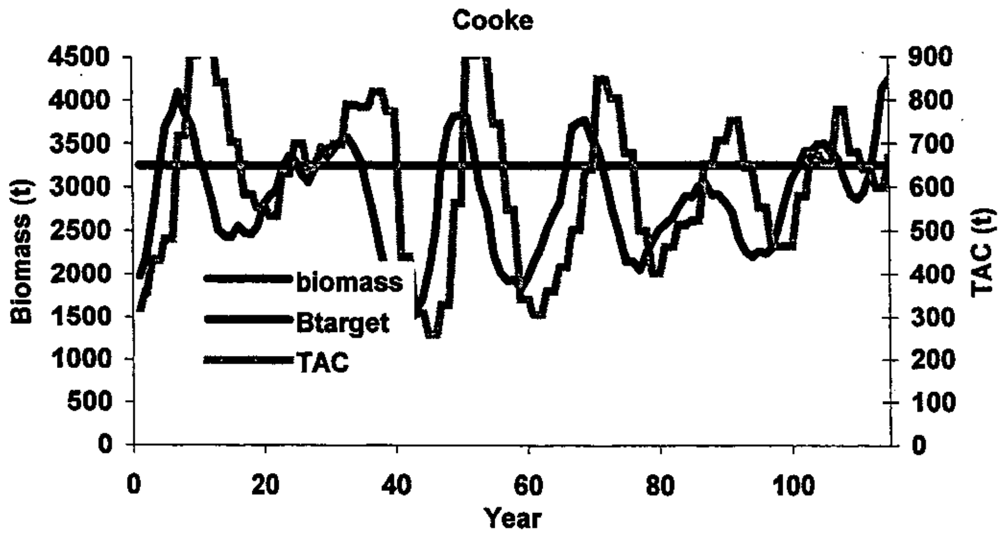

Figure 16: Biomass and TAC (grey) from a poor example of the Cooke rule, using a latent year.

5.1.5 Rule of Geromont et al. (1999) for Namibian hake

This rule performs much better without a latent year: the crash rate is an estimated 20% with a latent year. Setting the sensitivity parameter λ to 0.1 causes a very high crash rate even without a latent year; setting it to 0.8 seems to give higher fluctuations than intermediate values. We could detect no difference between $\lambda = 0.50$ and $\lambda = 0.35$, so chose $\lambda = 0.35$.

5.1.6 Rule of Fournier & Warburton (1988)

This rule shows very unstable behaviour (Figure 17) both with and without the latent year. We tried a wide variety of combinations of the three parameters without finding a region where the rule behaved reasonably. Reducing the size of the sensitivity parameter s reduces the size of fluctuations but disconnects TAC from biomass (Figure 18). We chose not to proceed further with this rule.

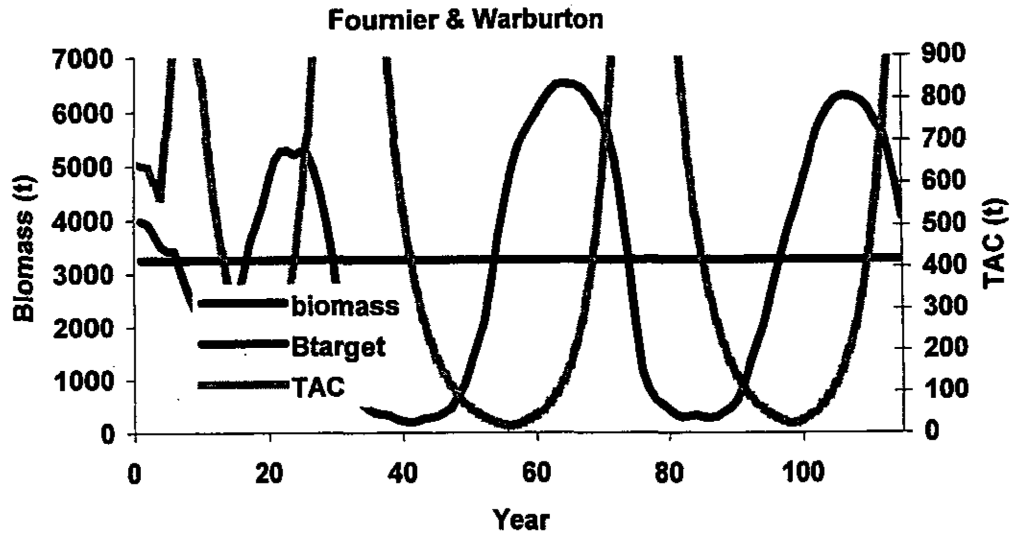

Figure 17: Behaviour of the Fournier & Warburton rule with no latent year and the parameters set by those authors.

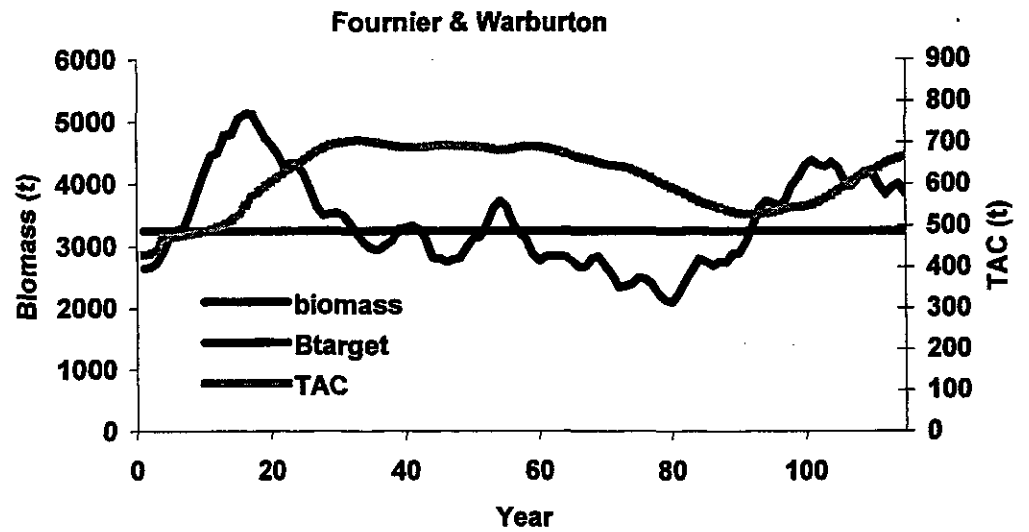

Figure 18: Behaviour of the Fournier & Warburton rule with no latent year and the sensitivity parameter reduced to 0.01.

5.1.7 Bentley's gradient rule

We left the threshold parameter T at zero for comparability with other rules, and adjusted the sensitivity parameter s to 0.01 after experimentation – higher values give increasing fluctuations. The rule is generally well behaved both with and without a latent year, but sometimes shows large fluctuations.

5.1.8 Rule of Bentley et al. (2002) for rock lobsters

This rule is not well behaved, with TAC lagging behind the biomass; adjusting the two parameters S and W appears to alter one type of poor behaviour to another. Figure 19 shows a typical example of the rule's behaviour with parameters set to those proposed for rebuilding the NSS. We chose not to proceed with this rule.

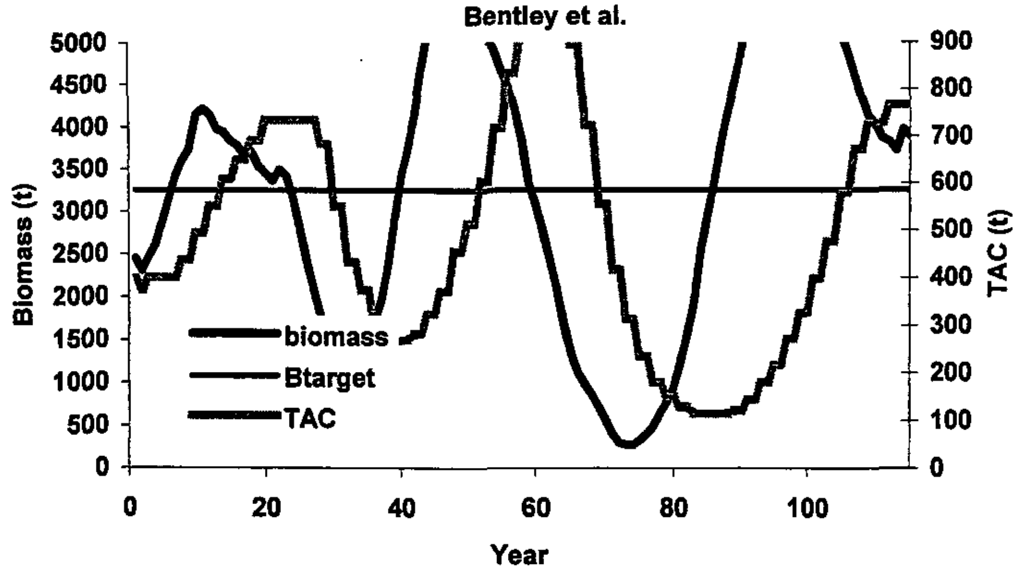

Figure 19: Behaviour of the Bentley et al. rule in a typical run with a latent year.

5.1.9 Rule of Haist (2002) for hoki

After experimentation, we set the parameters $\Delta = 0.9$ (smaller values give more fluctuations in biomass; higher values do not allow TAC to respond sufficiently); $s = 0.1$ (high values give poor behaviour) and minimum CPUE = 0.3 (the runs do not appear to go near this level).

With a latent year and parameters set as above, this rule gives *Bav* close to the target but sometimes shows long uncorrected fluctuation (Figure 20).

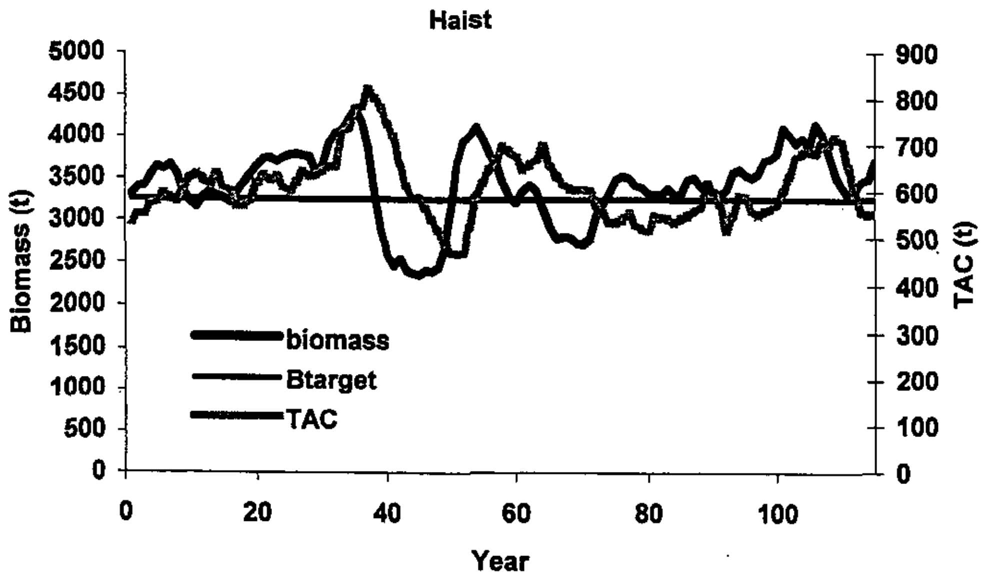

Figure 20: Behaviour of a random example of the Haist rule without a latent year.

5.1.10 Rule of Geromont & Glazer (1998) for South African hake

This is an F_n rule, and we used $n = 0.06$ to obtain a mean biomass near the target of 3250 t, as for the benchmark rule. We chose $\Delta = 0.1$ after observing its effect. The rule performs better without a latent year: large fluctuations occur with a latent year. This appears to be a generally well-behaved rule (Figure 21).

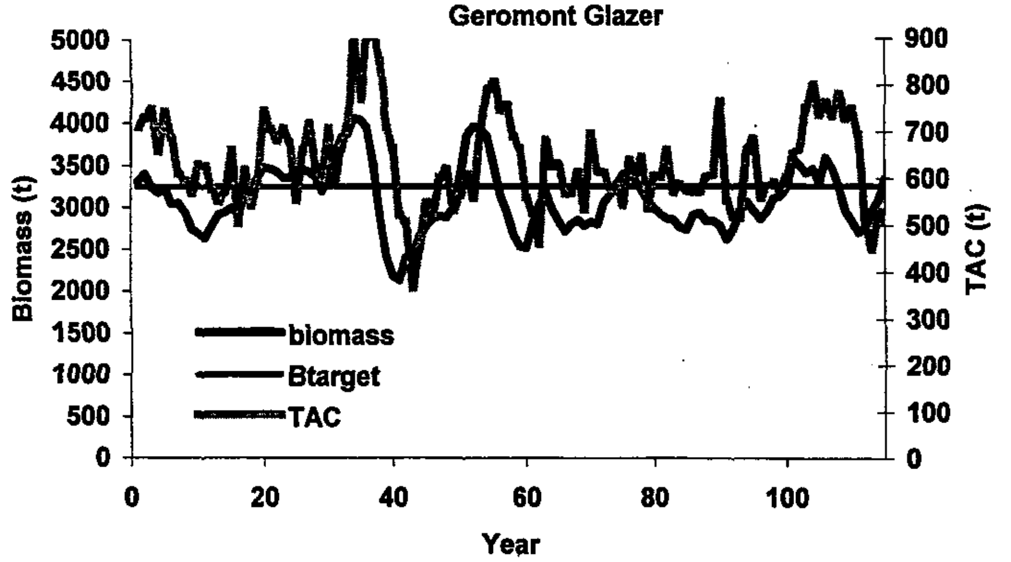

Figure 21: ***Behaviour of a random example of the Geromont & Glazer rule without a latent year.***

5.1.11 Bentley F_n rule

We used $n = 0.06$ again, and set $\Delta = 0.1$ as for the similar Geromont & Glazer's rule. Without a latent year, the rule appears to work generally well but shows more fluctuation than Geromont & Glazer (Figure 22). With a latent year the fluctuations in catch can become extreme, and have a tendency to get into cycles.

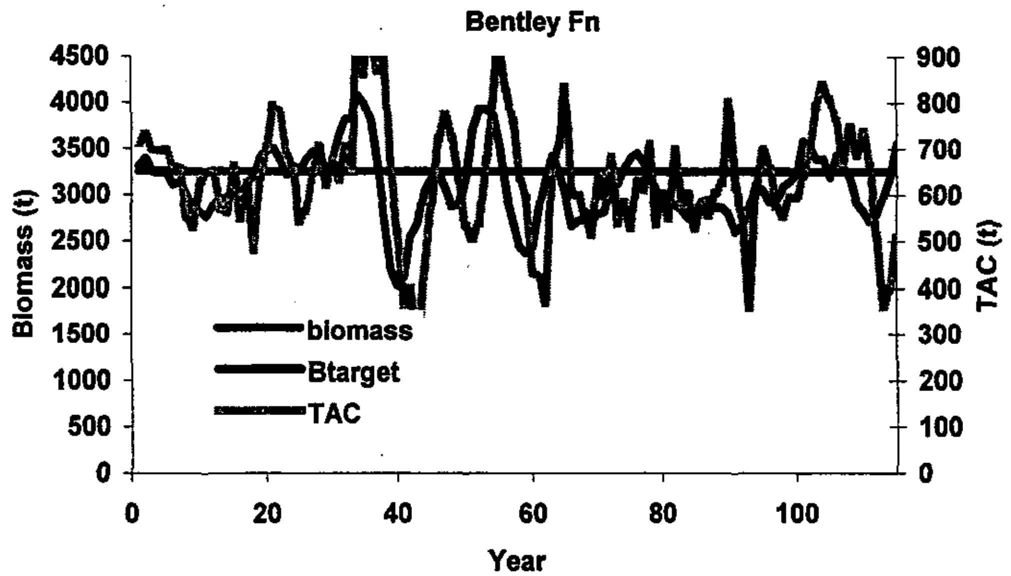

Figure 22: Behaviour of a random example of the Bentley F_n rule without a latent year.

5.1.12 Rule of Punt & Smith (1999) for gemfish

This rule has five parameters. We set $n = 0.06$ as for the other F_n rules, we set the minimum and maximum TACs at 50 and 800 respectively (large enough not to affect the rule much), and we set the minimum and maximum change to 0.75 and 1.5 respectively. With these settings, *Bav* is close to the target and we saw no crashes. A latent year increases fluctuations in catch and biomass. The rule appears to behave generally well (Figure 23) when compared with the benchmark rule.

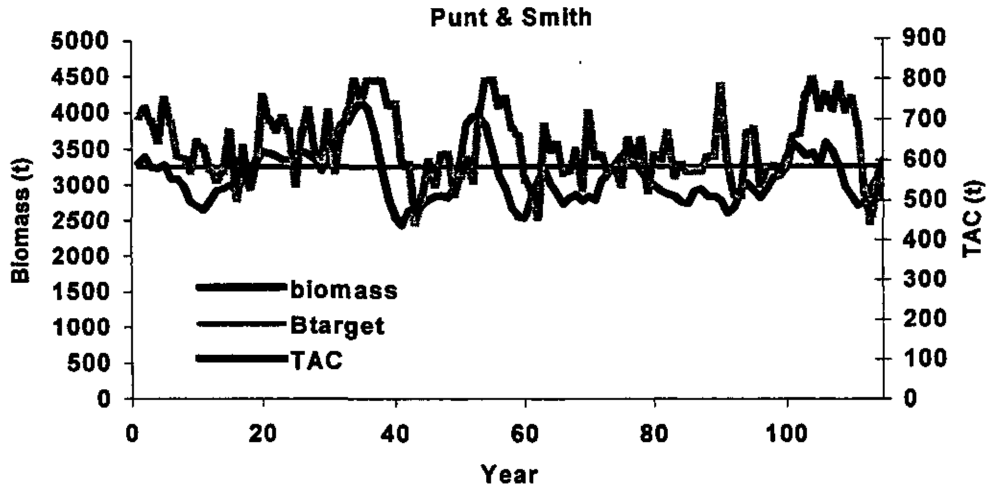

Figure 23: Behaviour of a random example of the Punt & Smith rule without a latent year.

5.1.13 Rule of De Oliveira et al. (1998) for sardines

This rule is the same as the Punt & Smith rule except for having one fewer parameter; we left the other parameters as above. Behaviour was very similar to that of the Punt & Smith rule.

5.1.14 Rule of Baldursson et al. (1996) for Icelandic cod

We left the parameters for this rule as described in Section 3.3.14 (see Figure 2). This rule tends to generate unstable cycles (Figure 24) and was not pursued further.

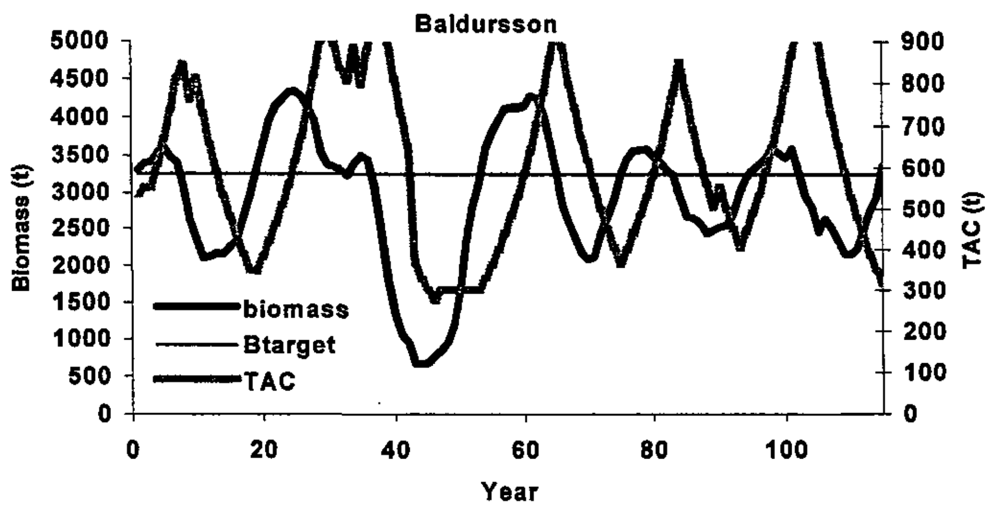

Figure 24: Behaviour of a random example of the Baldursson et al. rule without a latent year.

5.2 Simple comparison of ten rules

After discarding the rules of Bentley et al. (unpublished results), Fournier & Warburton (1988) and Baldursson et al. (1996), we made 1000 runs of the remaining rules. Table 5 shows the summary of results both with and without the latent year provision. We used the median of 1000 runs for the *Bav*, *Cav* and *AAV* indicators, the 20th percentile for the *Bmin*, *lowB* and *Cmin* indicators, and the 80th percentile for *Bdiff*.

Table 5: Showing results of 1000 runs of the remaining rules both with (lower) and without (upper) a latent year, using the parameters in Section 4.1. Grey cells indicate the worst performers and bold figures indicate the best performers in each trial for each indicator.

	median	20%	80%	median	80%	median	20%	
Without latent year								
Rule	***Bav***	***Bmin***	***lowB***	***AAV***	***Bdiff***	**Cav**	**Cmin**	**%Crash**
Benchmark	3312	**2590**	**4.29**	3.07	**1472**	**611**	**436**	0
Haist	3413	2049	12.99	3.13	2476	597	370	0
Geromont Glazer	3302	2159	16.02	10.35	1944	608	369	0
Geromont hake	3594	2161	14.71	**1.55**	2097	587	436	1.1
Punt & Smith	3355	2219	13.12	9.61	1936	605	373	0
De Oliveira	3308	2174	15.50	11.41	1921	607	367	0
Cooke	3237	1897	24.94	7.20	2246	608	294	0.1
Constant Catch	4389	3323	0.00	0.00	1744	500	398	1.8
Bentley gradient	3390	1973	23.10	4.95	2197	600	368	0
Bentley Fn	3299	2154	15.22	16.15	1947	607	272	0
simple	3481	1944	26.95	4.81	2099	593	367	0.3
With latent year								
Benchmark	3308	**2588**	**4.48**	3.07	**1493**	**610**	**435**	0
Haist	3505	1883	19.65	1.70	2815	585	387	4.3
Geromont Glazer	3277	2011	21.16	6.05	2095	606	358	0
Geromont hake	3796	1462	48.25	**0.76**	2709	541	314	14.6
Punt & Smith	3344	2133	15.89	5.36	2049	604	368	0
De Oliveira	3290	2052	19.85	16.45	2081	607	360	0
Cooke	3215	1700	32.85	16.42	2495	606	245	0
Constant catch	4385	3258	0.00	0.00	1800	499	399	1.8
Bentley gradient	3448	2089	16.98	4.00	2171	596	367	0.1
Bentley Fn	3279	1969	21.67	19.70	2173	607	271	0
simple	3637	1100	80.64	2.30	3161	560	308	10.3

Mean biomass, *Bav*: all the rules do reasonably well at producing a mean biomass near the target except the constant catch rule with a TAC of 500 t. A higher TAC for this rule would come closer to target mean biomass at the cost of increased crash frequency (see Figure 12).

Minimum biomass, *Bmin*: we looked at the 20th percentiles for this indicator. The worst rules with the latent year are the simple, Cooke, and Geromont hake; and without the latent year are Cooke, simple and Bentley gradient rules, all performing better than they did with a latent year. The best rule (ignoring constant catch) is the benchmark.

Low biomass index, *lowB*: Without a latent year, the worst rules are Haist, Cooke and the benchmark. With a latent year, the worst rules are Cooke, Bentley F_n and Haist. The best rule is Geromont hake.

Average annual variation in catch, *AAV*: Without a latent year, the worst rules are Bentley F_n, De Oliveira and Geromont & Glazer. With a latent year, the worst rules are Bentley F_n, De Oliveira and Cooke. The best rule is Geromont hake.

Biomass range, *Bdiff*: Without a latent year, the worst rules are Haist, Cooke and Bentley gradient rules. With a latent year, the worst rules are simple, Haist and Geromont hake. The best rule is the benchmark.

Mean catch, *Cav*: Without a latent year, the worst rules are Geromont hake, simple and Haist. With a latent year, the worst rules are Geromont hake, simple and Haist. The best rule is the benchmark.

Minimum catch, *Cmin*: Without a latent year, the worst rules are Bentley F_n, Cooke, simple and De Oliveira. With a latent year, the worst rules are Cooke, Bentley F_n and simple. The best rule is the benchmark.

Percentage of crashes, *%Crash*: Most rules do not crash without a latent year (zero to 1% crash rate). The Geromont hake rule crashed 11 times out of 1000 runs and the constant catch rule 18 times. With a latent year, the Geromont hake and simple rules crash for more than 10% of runs (more crashes than the constant catch rule) and Haist rule crashed 13 times out of 1000 runs. The benchmark, Geromont & Glazer, Punt & Smith, De Oliveira and Bentley F_n rules did not crash in these trials.

In summary, most rules performed better (except for *AAV*) without a latent year, and some were unacceptable with a latent year because of the high crash rate. The constant catch rule delivered a much lower catch than the other rules, with a 2% crash rate. The different rules show substantial differences in some indicators – *Bmin*, *lowB*, *AAV* and *Bdiff* – but not much difference in *Bav* and *Cav*.

5.3 Fine tuning selected rules

For some rules we experimented further with their parameter sets to find parameters that give good behaviour. We did this with the Haist, Cooke, Bentley gradient, Bentley F_n, Punt & Smith and Geromont & Glazer rules. The value of Δ was looked at closely for Haist, Bentley's F_n and Geromont & Glazer rules; parameters a and b were investigated for Cooke's rule, and the threshold, T, and sensitivity, s, for Bentley's gradient rule were explored. For the Haist rule we also looked at the parameter s. We fixed $n = 0.06$ for all rules using F_n. Tuning of parameters for each rule was done with the latent year because most rules behave worse with than without the latent year, and because the stakeholders indicate a strong preference for a latent year.

5.3.1 Cooke's rule

Table 6 shows the value for the indicators when various parameter sets were used for the Cooke rule.

As a increases, mean biomass increases and mean catch decreases. We chose $a = -200$ as a trade-off between biomass and catch. As b increases, the mean biomass and minimum catch decrease but mean catch increases. We chose $b = 0.22$ to have large minimum and average catches with relatively large biomass (Table 6). The tuning of Cooke's rule led to the better indicators than those in section 7.2.

Table 6: Summary of indicators from 1000 runs with various parameter sets for the Cooke rule. The shaded parameter set is the final choice.

a	*b*	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
-100	0.15	4187	3158	0.01	4.00	1711	531	348	0.0
-100	0.25	2870	1248	69.96	6.41	2754	613	225	0.5
-100	0.35	1906	270	321.23	22.03	3555	553	7	15.6
-190	0.25	3197	1673	33.57	6.43	2508	606	251	0.0
-200	0.15	4537	3586	0.00	4.51	1501	485	310	0.0
-200	0.2	3831	2653	1.88	5.17	1944	568	328	0.0
-200	0.22	3559	2260	8.33	5.54	2154	587	307	0.0
-200	0.24	3319	1905	21.73	6.11	2378	601	282	0.0
-200	0.25	3242	1695	31.74	6.45	2507	605	247	0.0
-200	0.26	3103	1504	43.66	6.91	2687	609	213	0.0
-200	0.35	2262	332	252.04	13.22	3807	582	0	5.5
-210	0.25	3262	1808	27.99	6.42	2407	603	256	0.1
-300	0.15	4888	3977	0.00	5.26	1358	434	264	0.0
-300	0.25	3517	2096	11.08	6.70	2360	588	258	0.0
-300	0.35	2582	407	182.08	11.39	3871	599	0	1.8

5.3.2 Bentley's gradient rule

Table 7 shows the values for the indicators when various parameter sets were used for Bentley's gradient rule.

Table 7: Summary of indicators from 1000 runs with various parameter sets for Bentley's gradient rule. The shaded parameter set is the final choice.

T	*s*	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
0	0.007	3494	2095	15.72	3.95	2170	593	369	0
0	0.01	3474	2043	16.68	3.99	2217	597	365	0
0	0.05	3317	1949	19.40	4.33	2400	605	346	0
0	0.2	3240	1423	32.91	5.71	3133	600	237	0
0.1	0.01	3463	2076	17.12	3.96	2193	595	368	0
0.3	0.01	3460	2096	17.57	4.02	2156	597	370	0

As T increases, mean biomass decreases and minimum catch increases. Mean catch does not change with increasing T. We chose $T = 0$. As s increases, mean biomass and minimum catch decrease, but mean catch increases. We chose $s = 0.01$ to have large minimum and average catches with relatively large biomass (Table 7).

5.3.3 Haist's rule

Table 8 shows indicator values when various parameter sets were used for the Haist rule. The rule was sensitive to both parameters tested.

Table 8: Summary of indicators from 1000 runs with various parameter sets of Haist. The shaded parameter set is the final choice.

s	Δ	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
0.09	0.9	3534	1866	19.85	1.67	2843	580	383	1.2
0.1	0.8	3448	1821	19.12	3.47	2898	588	331	0.2
0.1	0.86	3465	1853	19.09	2.37	2827	585	362	0.6
0.1	0.88	3489	1867	18.99	2.01	2772	586	374	0.7
0.1	0.89	3477	1855	20.81	1.85	2816	586	381	1.4
0.1	0.9	3514	1827	20.74	1.70	2864	583	384	1.6
0.1	0.91	3523	1874	19.36	1.53	2796	582	389	1.8
0.11	0.86	3431	1753	22.38	2.40	2925	590	361	0.9
0.11	0.88	3454	1820	20.86	2.05	2874	586	374	1.7
0.11	0.9	3462	1815	21.03	1.69	2848	587	387	1.8
0.12	0.86	3417	1778	22.35	2.43	2868	591	364	0.7
0.12	0.88	3423	1737	24.89	2.09	2963	590	369	1.8
0.12	0.9	3442	1732	25.06	1.73	2960	587	380	1.8
0.2	0.9	3275	1278	47.74	1.93	3472	595	339	5.5

As s increases, mean biomass decreases and mean catch increases. We chose $s = 0.1$ as a trade-off between biomass and catch. As Δ increases, mean biomass and minimum catch increase but mean catch decreases. We chose $\Delta = 0.9$ to have large minimum and average catches with relatively large biomass (Table 8).

5.3.4 Geromont & Glazer's rule

Table 9 and Figure 25 show the values for the indicators when various parameter sets were used for the Geromont & Glazer rule. When Δ is big (0.9) the crash rate increases to 30% with poor values for all indicators except the *AAV* (Table 9).

Table 9: Summary of indicators from 1000 runs with various parameter sets for the Geromont & Glazer rule. The shaded parameter set is the final choice.

delta	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
0.1	3297	2034	20.08	6.05	2081	606	359	0.0
0.2	3279	2033	19.16	5.25	2133	607	362	0.0
0.3	3294	2012	18.72	4.63	2167	605	369	0.0
0.4	3291	1956	22.47	3.97	2224	606	372	0.1
0.5	3288	1916	24.98	3.34	2303	605	376	0.2
0.6	3297	1807	26.59	2.76	2450	600	370	0.3
0.7	3272	1683	32.52	2.15	2640	599	370	1.4
0.8	3243	1146	61.36	1.61	3293	593	327	8.2
0.9	3060	199	267.40	1.13	4831	576	146	30.0

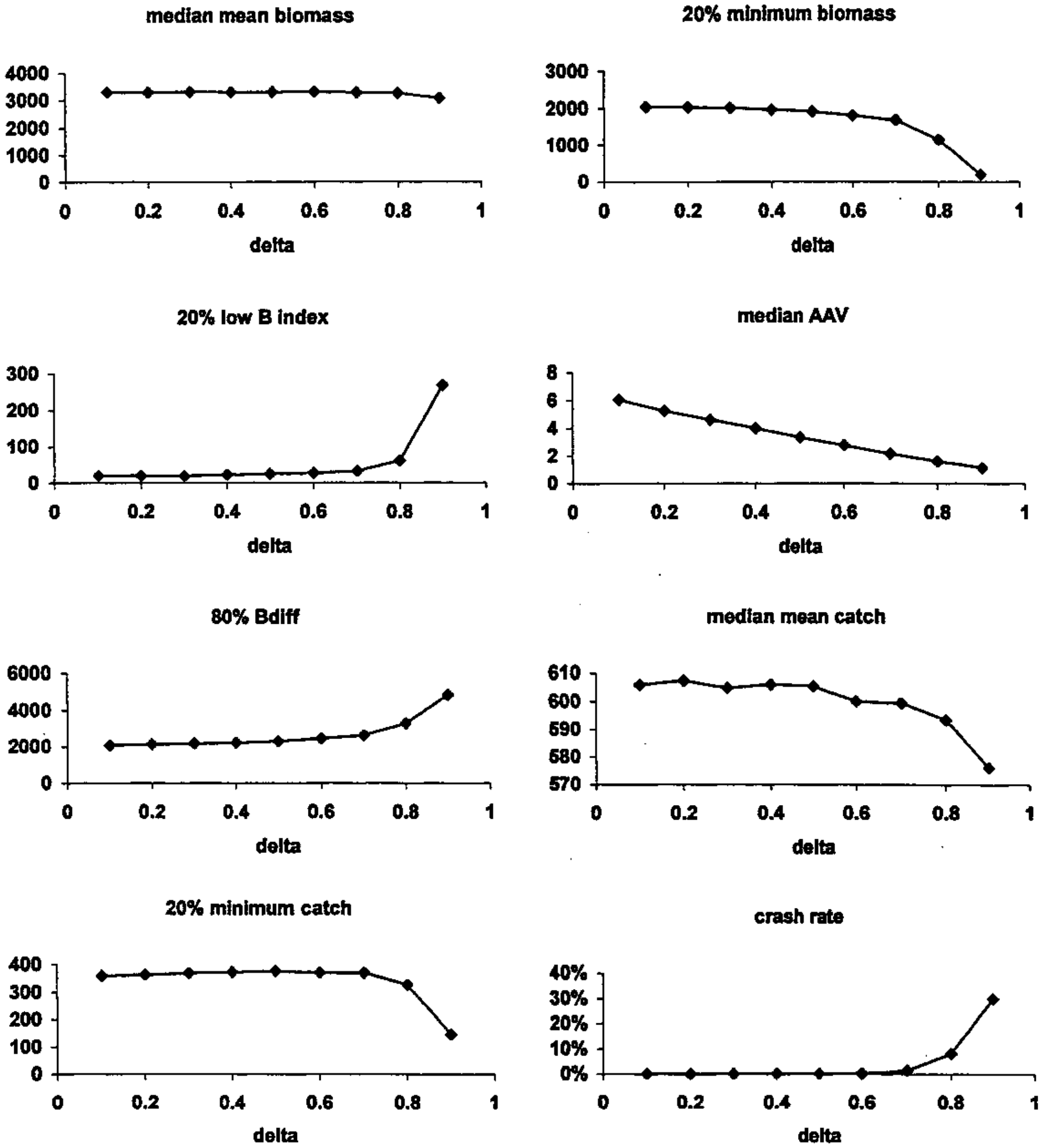

Figure 25: Summary indicators from 1000 runs with various parameter sets for the Geromont & Glazer rule.

Figure 25 shows insensitivity to Δ below Δ = 0.7 for most indicators, then increasing sensitivity from Δ = 0.7 to Δ = 0.8. We decided to use Δ = 0.4 because both catch indicators are higher than when other Δ are used, and the crash rate is lower than Δ = 0.5.

5.3.5 Bentley's F_n rule

Table 10 shows the values for the indicators when various parameter sets were used for Bentley's F_n rule. When Δ is big (0.9) the crash rate increased to 12.6% with poor values for *Bmin*, *lowB*, *Bdiff* and *Cav* (Table 10). The crash rate began to increase from zero at Δ=0.6. We decided to use Δ=0.5 because both catch indicators are reasonably high, *Bav* is at its highest, the crash rate is zero and the other indicators are acceptable.

Table 10: Summary of indicators from 1000 runs with various parameter sets for the Bentley F_n rule. The shaded parameter set is the final choice.

Δ	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
0.1	3287	1970	21.67	8.42	2198	606	291	0.0
0.2	3284	1927	21.96	7.35	2199	607	306	0.0
0.3	3300	2004	20.08	6.27	2215	606	330	0.0
0.4	3291	1941	21.45	5.28	2229	606	342	0.0
0.5	3296	1925	22.70	4.34	2275	605	356	0.0
0.6	3295	1879	24.27	3.44	2346	604	372	0.1
0.7	3281	1834	26.81	2.60	2408	604	389	0.6
0.8	3269	1687	33.85	1.81	2615	599	393	1.4
0.9	3209	997	86.29	1.02	3473	592	361	12.6

5.3.6 Punt & Smith's rule

Table 11 shows the values for the indicators when various parameter sets were used for Punt & Smith rule.

Table 11: Summary of indicators from 1000 runs with various parameter sets for Punt & Smith rule. The shaded parameter set is the final choice.

TAC_{min}	TAC_{max}	$Max_{decrease}$	$Max_{increase}$	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
25	800	0.75	1.5	3299	2058	18.25	6.42	2085	606	363	0.0
50	600	0.75	1.5	3641	2463	4.68	1.92	2028	582	397	0.0
50	700	0.75	1.5	3358	2141	14.66	5.30	2064	604	367	0.0
50	800	0.75	1.5	3303	2066	17.80	6.42	2085	606	357	0.0
50	900	0.75	1.5	3289	2071	18.39	6.63	2113	607	358	0.0
200	800	0.75	1.5	3305	2087	17.91	6.37	2085	607	360	0.0
50	800	0.25	1.25	3311	2103	16.86	6.16	2098	604	354	0.0
50	800	0.25	1.5	3290	2077	18.19	6.54	2077	606	356	0.0
50	800	0.25	1.75	3299	2061	18.52	6.54	2101	607	353	0.1
50	800	0.5	1.25	3326	2079	17.32	6.17	2107	607	354	0.0
50	800	0.5	1.5	3288	2086	18.27	6.50	2087	606	358	0.0
50	800	0.5	1.75	3311	2092	17.08	6.55	2063	607	352	0.0
50	800	0.75	1.25	3318	2096	16.08	6.00	2095	605	361	0.1
50	800	0.75	1.75	3292	2083	18.17	6.51	2075	607	356	0.0

As TAC_{max} increases, the average biomass decreased towards the target CPUE. Changing TAC_{min} does not change any indicators (Table 11). Larger $Max_{increase}$, gives larger *lowB* and *AAV*. Larger $Max_{decrease}$ gives larger *lowB* and *AAV*, except when $Max_{increase}$ was 1.25, where $Max_{decrease}$ = 0.75 gives the lowest lowBindex and AAV. We chose the shaded parameter set (Table 11) because it had the lowest *lowB* and *AAV* with reasonable mean catch and biomass.

5.3.7 Comparison of tuned rules

Table 12 summarises results from the tuned harvest control rules compared to the benchmark. The average biomass is close to target biomass for all rules, the 20% minimum biomass is less than the benchmark (1809 – 2281 *vs* 2487) and the low biomass index is sometimes higher (Haist, Bentley's Fn, Punt & Smith and Geromont & Glazer) or sometimes lower (Cooke and Bentley's gradient rule) than the benchmark. The average annual variation is always larger than the benchmark except for the Haist rule and the Punt & Smith and Cooke rule had much higher AAV then the benchmark. The difference between the maximum and the minimum biomass is always higher than the benchmark. The average catch is similar for all rules, but always less than the benchmark, and the minimum catch is smaller with the Cooke's rule compared to other rules, but not bad. The crash rate is very low for all rules, with Haist's rule having the highest crash rate of 2.0%.

Figure 26 shows the behaviour of all tuned rules and the pattern of production deviations from a typical run. The Cooke and Bentley gradient rules, and to a lesser extent the Geromont & Glazer rule, show fluctuations that are unrelated to the production deviations and may be induced by the implementation lag in the rules. The Haist rule gives a catch trajectory that seems disconnected from the biomass trajectory.

Table 12: Summary of indicators from 1000 runs with tuned harvest control rules and the benchmark rule. The upper part shows the actual indicator values, and the lower part shows the proportionate difference between the rules indicated and the benchmark rule (negative indicates worse, positive better)

	median	20%	80%	median	80%	median	20%	crash
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	Cav	Cmin	rate (%)
Benchmark	3312	2487	6.03	3.27	1615	610	435	0.0
Cooke	3571	2281	8.53	5.55	2141	588	311	0.0
Bentley's gradient	3444	2105	15.90	4.01	2161	597	368	0.0
Haist	3499	1809	20.96	1.69	2860	584	384	2.0
Geromont & Glazer	3281	1962	22.62	3.94	2216	606	374	0.0
Bentley Fn	3289	1906	23.93	4.33	2260	607	361	0.0
Punt & Smith	3310	2087	17.86	6.05	2127	607	360	0.0

	median	20%	80%	median	80%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	Cav	Cmin
Cooke	0.08	-0.08	[redacted]		-0.33	-0.04	-0.29
Bentley's gradient	0.04	-0.15	[redacted]	-0.22	-0.34	-0.02	-0.15
Haist	0.06	-0.27	-2.48	0.48	[redacted]	-0.04	-0.12
Geromont & Glazer	-0.01	-0.21	-2.75	-0.20	-0.37	-0.01	-0.14
Bentley F_n	-0.01	-0.23	-2.97	-0.32	-0.40	-0.01	-0.17
Punt & Smith	0.00	-0.16	-1.96	[redacted]	-0.32	-0.01	-0.17

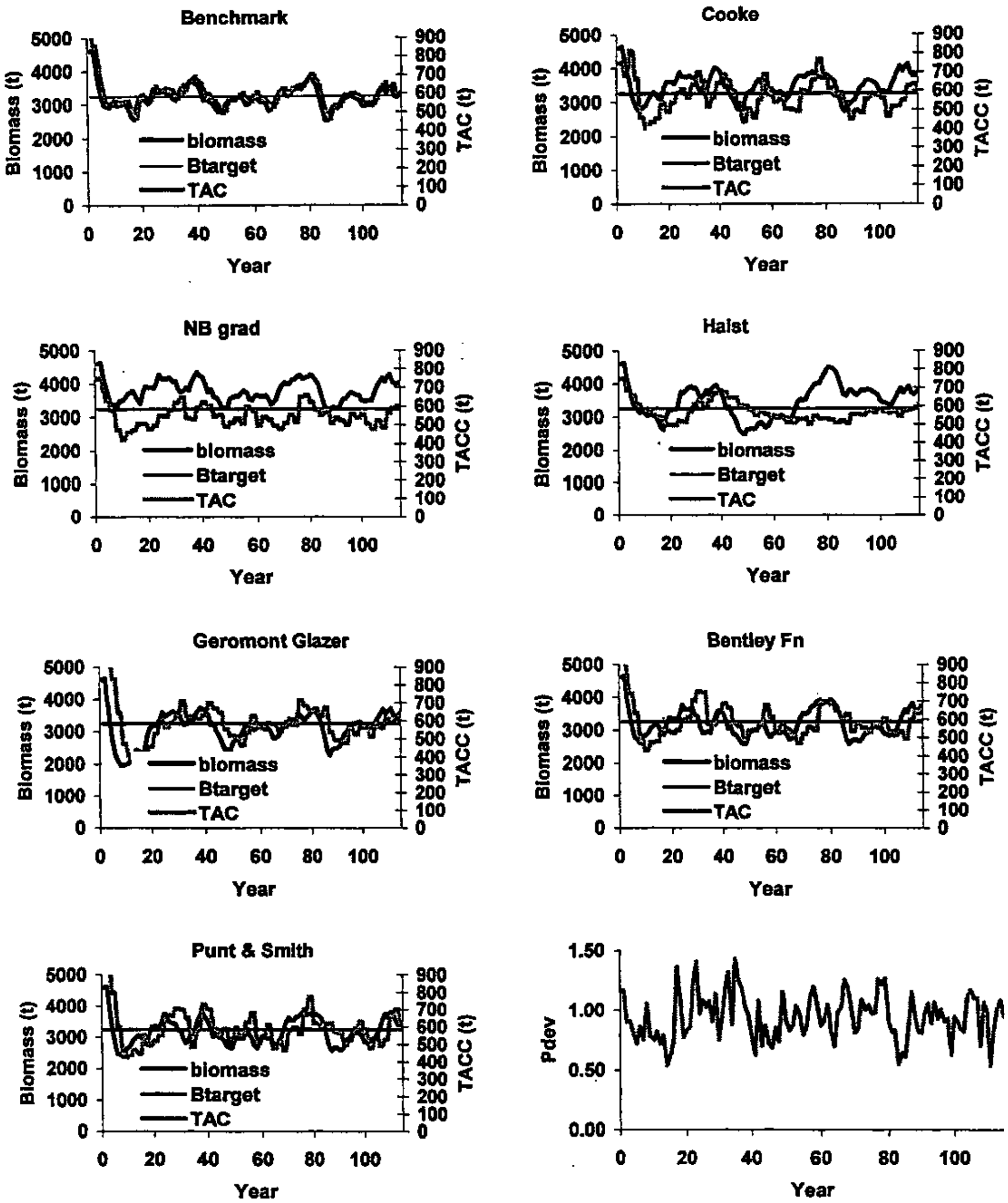

Figure 26: Tuned rules and the production deviations (bottom right).

5.4 Robustness testing

5.4.1 Effect of robustness trials on the benchmark rule

Table 13 shows the indicators from the benchmark rule only for each of the robustness trials. Most indicators became worse under most robustness trials. The most sensitive indicators were *Bdiff*, *Bmin* and *lowB*. The least sensitive indicators were *Cav* and *Bav*. No trial other than r9c and r9d caused a measurable crash rate.

Episodic mortality had a large negative effect on the biomass indicators, obviously because it decreased biomass at random intervals. Despite this, median biomass and catch decreased by only 5%. Increasing *q* over the run also had a large negative effect, because CPUE becomes increasingly misleading during the run, causing the rule to take the catch from a smaller than desired biomass.

Varying the population parameters randomly had no effect on the median biomass, catch or *AAV*, but the distributions of *Bmin* and *Cmin* became wider. The introduced production lag had a very small effect except that the low biomass index became worse. Varying *K* also had little effect on most indicators, but a large effect on the *lowB* index. Unreported catch had a surprisingly small effect on the results.

Uncontrolled catch had large effects, depending on its amount relative to the commercial catch and especially on whether it increased or not. When the uncontrolled catch was in balance with the commercial catch (trial r9a), there was little effect on any indicator. A large uncontrolled catch caused all indicators to deteriorate except total catch. When uncontrolled catch was allowed to increase, even by a small annual amount, substantial crash rates were seen and almost all indicators decreased dramatically.

Table 13: Effects of the various robustness trials on the indicators from the benchmark rule. The top part shows the actual indicator values, and the lower part shows the percentage difference (negative indicates worse, positive indicates better) between the base case and the robustness trial shown.

	code	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)	median *Clav*	20% *Clmin*
	base	3312	2487	6.0	3.27	1615	610	435	0.0		
parameters vary	r1	3050	1840	67.2	3.31	1664	563	326	0.0		
production lag	r2	3311	2478	6.4	3.46	1769	610	433	0.0		
varying *K*	r3	3309	1682	46.7	3.61	3658	610	294	0.0		
regime shift	r4	3336	2135	30.0	3.33	2236	613	369	0.0		
episodic mortality	r5	3156	1496	29.9	4.98	2591	581	270	0.0		
doubled error,											
σ_ε, σ_δ and σ_θ	r6a	3240	1700	30.9	6.61	3215	596	287	0.0		
σ_ε only	r6b	3248	1881	24.4	6.25	3033	600	335	0.0		
σ_δ only	r6c	3310	2245	10.7	3.81	2006	607	360	0.0		
σ_θ only	r6d	3323	2495	5.7	3.26	1629	610	436	0.0		
increasing *q*	r7	2544	1359	88.2	4.64	2349	639	399	0.0		
unreported catch	r8	3073	2170	22.0	3.63	1660	621	430	0.0		
uncontrolled catch,											
α=75, %increase =0	r9a	3304	2531	5.4	3.07	1533	612	470	0.0	306	218
α=200, %increase=0	r9b	2655	1726	64.4	3.98	1816	630	466	0.0	245	152
α=75, %increase=1	r9c	2108	245	232.6	5.44	3286	610	184	21.7	197	14
α=200, %increase=1	r9d	897	133	673.2	9.65	2832	394	99	99.5	76	2

	code	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin
parameters vary	r1	-0.08	-0.26	-10.16	-0.01	-0.03	-0.08	-0.25
production lag	r2	0.00	0.00	-0.07	-0.06	-0.10	0.00	-0.01
varying *K*	r3	0.00	-0.32	-6.75	-0.10	-1.26	0.00	-0.32
regime shift	r4	0.01	-0.14	-3.98	-0.02	-0.38	0.00	-0.15
episodic mortality	r5	-0.05	-0.40	-3.96	-0.52	-0.60	-0.05	-0.38
doubled error,								
σ_ε, σ_δ and σ_θ	r6a	-0.02	-0.32	-4.12	-1.02	-0.99	-0.02	-0.34
σ_ε only	r6b	-0.02	-0.24	-3.04	-0.91	-0.88	-0.02	-0.23
σ_δ only	r6c	0.00	-0.10	-0.78	-0.16	-0.24	-0.01	-0.17
σ_θ only	r6d	0.00	0.00	0.06	0.00	-0.01	0.00	0.00
increasing *q*	r7	-0.23	-0.45	-13.63	-0.42	-0.45	0.05	-0.08
unreported catch	r8	-0.07	-0.13	-2.65	-0.11	-0.03	0.02	-0.01
uncontrolled catch,								
α=75, %increase =0	r9a	0.00	0.02	0.10	0.06	0.05	0.00	0.08
α=200, %increase=0	r9b	-0.20	-0.31	-9.69	-0.22	-0.12	0.03	0.07
α=75, %increase=1	r9c	-0.36	-0.90	-37.60	-0.66	-1.03	0.00	-0.58
α=200, %increase=1	r9d	-0.73	-0.95	-110.72	-1.95	-0.75	-0.35	-0.77

5.4.2 Effects of robustness trials on the selected rules

5.4.2.1 Varying production parameters

Results from trial r1, where the population model parameters were varied randomly between runs, are shown in Table 14. Only Cooke's and Bentley's gradient rules were free of crashes, Haist had 6% and the others had smaller rates from 2.5%.

All rules performed worse than the benchmark (Table 14b) for all three catch indicators, except that Haist's rule performed substantially better for *AAV* (and *lowB*). In the other rules there was little difference in *Bav*. Three rules: Cooke's, Bentley's gradient and Haist's, performed better with respect to low biomass indicators than did the benchmark, while the other three performed much worse. This trial thus showed some contrast among the various rules. Differences between these results and the base case results (Table 14c) were mainly in the *Cmin* and low biomass indicators, where all were worse except for the Haist rule's *lowB*, which was better than in the base case.

Table 14a: Summary of indicators from 1000 runs, and the lower part shows the difference between the rules indicated and the benchmark rule (negative indicates worse, positive better) in robustness trial r1, showing the actual indicator values.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3050	1840	67.25	3.31	1664	563	326	0.0
Cooke	3383	1920	27.83	5.85	2209	543	219	0.1
Bentley's gradient	3338	1955	28.35	3.89	2183	544	290	0.0
Haist	3758	1828	20.35	1.92	2958	506	271	5.9
Geromont & Glazer	3000	1306	102.29	4.18	2381	554	247	4.5
Bentley F_n	3007	1304	100.63	4.54	2382	554	239	2.5
Punt & Smith	3045	1517	84.27	6.10	2159	555	255	2.5

Table 14b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.11	0.04	0.59	-0.77	-0.33	-0.04	-0.33
Bentley's gradient	0.09	0.06	0.58	-0.18	-0.31	-0.03	-0.11
Haist	0.23	-0.01	0.70	0.42	-0.78	-0.10	-0.17
Geromont & Glazer	-0.02	-0.29	-0.52	-0.26	-0.43	-0.02	-0.24
Bentley F_n	-0.01	-0.29	-0.50	-0.37	-0.43	-0.02	-0.27
Punt & Smith	0.00	-0.18	-0.25	-0.84	-0.30	-0.01	-0.22

Table 14c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	-0.08	-0.26	-10.16	-0.01	-0.03	-0.08	-0.25
Cooke	-0.05	-0.16	-2.26	-0.05	-0.03	-0.08	-0.29
Bentley's gradient	-0.03	-0.07	-0.78	0.03	-0.01	-0.09	-0.21
Haist	0.07	0.01	0.03	-0.14	-0.03	-0.13	-0.30
Geromont & Glazer	-0.09	-0.33	-3.52	-0.06	-0.07	-0.09	-0.34
Bentley Fn	-0.09	-0.32	-3.21	-0.05	-0.05	-0.09	-0.34
Punt & Smith	-0.08	-0.27	-3.72	-0.01	-0.01	-0.08	-0.29

5.4.2.2 Lag in productivity

Table 15 shows results from trial r2, in which productivity was lagged to the biomass two years previously.

The only crashes seen were 0.2% in Haist's rule. All rules performed worse than the benchmark for all three catch indicators, except that Haist's rule performed substantially better for *AAV* than did the benchmark. The difference in *Cav* was slight for all rules. All rules were similar to the benchmark in *Bav*.

The main difference among rules was that Cooke's and Bentley's gradient rules had better *lowB* than the benchmark rule while all the others were worse.

Compared with the base case results, there was little change except that *lowB* and *Bdiff* were reduced for all rules, but the differences were not great (*see* Table 12).

Table 15a: Summary of indicators from 1000 runs in robustness trial r2.

	median	20%	80%	median	80%	median	20%	crash
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*Cav*	*Cmin*	rate (%)
Benchmark	3311	2478	6.44	3.46	1769	610	433	0.0
Cooke	3558	2094	12.34	6.38	2640	585	277	0.0
Bentley's gradient	3437	1988	20.65	4.40	2516	595	354	0.0
Haist	3522	1878	19.01	1.75	3021	581	383	0.2
Geromont & Glazer	3278	1957	24.14	4.20	2476	604	367	0.0
Bentley F_n	3274	1895	25.36	4.66	2542	604	355	0.0
Punt & Smith	3306	2069	18.78	6.17	2338	605	356	0.0

Table 15b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median	20%	80%	median	80%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*Cav*	*Cmin*
Cooke	0.07	-0.16	-0.91	-0.85	-0.49	-0.04	-0.36
Bentley's gradient	0.04	-0.20	-2.20	-0.27	-0.42	-0.02	-0.18
Haist	0.06	-0.24	-1.95	0.49	-0.71	-0.05	-0.11
Geromont & Glazer	-0.01	-0.21	-2.75	-0.22	-0.40	-0.01	-0.15
Bentley F_n	-0.01	-0.24	-2.94	-0.35	-0.44	-0.01	-0.18
Punt & Smith	0.00	-0.17	-1.91	-0.79	-0.32	-0.01	-0.18

Table 15c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median	20%	80%	median	80%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*Cav*	*Cmin*
Benchmark	0.00	0.00	-0.07	-0.06	-0.10	0.00	-0.01
Cooke	0.00	-0.08	-0.45	-0.15	-0.23	-0.01	-0.11
Bentley's gradient	0.00	-0.06	-0.30	-0.10	-0.16	0.00	-0.04
Haist	0.01	0.04	0.09	-0.04	-0.06	0.00	0.00
Geromont & Glazer	0.00	0.00	-0.07	-0.07	-0.12	0.00	-0.02
Bentley Fn	0.00	-0.01	-0.06	-0.07	-0.13	0.00	-0.02
Punt & Smith	0.00	-0.01	-0.05	-0.02	-0.10	0.00	-0.01

5.4.2.3 Variable productivity

Table 16 shows results from trial r3, in which *K* varied through one cycle of a cosine curve.

The low biomass index was very high for all rules, compared with the base case (Table 12). Some rules showed a slight crash rate, less than 1%, but Haist's rule had a 40% crash rate. Although median *Bav* was high, the range of biomass *Bdiff* was very large.

Median *Bav* did not change much (except in Haist's rule), *Bmin* was lower than the benchmark's for all rules except Cooke's, median *AAV* was also worse than the benchmark for all rules but Haist's. For all rules *Cmin* decreased compared with the benchmark but *Cav* was little changed (except in Haist's rule).

Compared with the base case results, all rules showed much worse *lowB* index, greater *Bdiff* and lower *Cmin*. *Bav*, *Cav* and *AAV* were generally unaffected.

Table 16: Summary of indicators from 1000 runs in robustness trial r3.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3309	1682	46.69	3.61	3658	610	294	0.0
Cooke	3561	1825	30.59	5.98	3821	585	202	0.0
Bentley's gradient	3339	1309	72.68	4.51	4207	603	242	0.7
Haist	4169	132	292.31	3.41	7672	451	88	39.4
Geromont & Glazer	3279	1372	67.97	4.31	4148	607	255	0.3
Bentley F_n	3285	1351	66.83	4.65	4157	606	246	0.1
Punt & Smith	3465	1460	61.06	4.85	4412	595	247	0.2

Table 16b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.08	0.09	0.34	-0.66	-0.04	-0.04	-0.31
Bentley's gradient	0.01	-0.22	-0.56	-0.25	-0.15	-0.01	-0.18
Haist	0.26	-0.92	-5.26	0.06	-1.10	-0.26	-0.70
Geromont & Glazer	-0.01	-0.18	-0.46	-0.20	-0.13	-0.01	-0.13
Bentley F_n	-0.01	-0.20	-0.43	-0.29	-0.14	-0.01	-0.16
Punt & Smith	0.05	-0.13	-0.31	-0.34	-0.21	-0.02	-0.16

Table 16c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	0.00	-0.32	-6.75	-0.10	-1.26	0.00	-0.32
Cooke	0.00	-0.20	-2.59	-0.08	-0.78	-0.01	-0.35
Bentley's gradient	-0.03	-0.38	-3.57	-0.13	-0.95	0.01	-0.34
Haist	0.19	-0.93	-12.95	-1.02	-1.68	-0.23	-0.77
Geromont & Glazer	0.00	-0.30	-2.00	-0.09	-0.87	0.00	-0.32
Bentley Fn	0.00	-0.29	-1.79	-0.07	-0.84	0.00	-0.32
Punt & Smith	0.05	-0.30	-2.42	0.20	-1.07	-0.02	-0.31

5.4.2.4 Regime shift

Table 17 summarises the results from trial r4, where regime shifts in *K* were simulated.

Most rules had crash rates less than 1%, but Haist's rule was 7%. The average indicators were little different from the benchmark or the base case (because the regime shift produced a new *K* with a mean the same as the old *K*), but the *Bmin* and *Cmin* indicators were generally worse than the benchmark rule and base case. The *lowB* index was worse than the benchmark and base case for some rules and better in others, without much pattern.

Table 17a: Summary of indicators from 1000 runs in robustness trial r4.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3336	2135	29.99	3.33	2236	613	369	0.0
Cooke	3547	2042	19.54	5.73	2704	588	257	0.3
Bentley's gradient	3399	1776	37.85	4.15	2745	600	305	0.6
Haist	3492	1362	43.72	1.82	3799	576	277	6.9
Geromont & Glazer	3258	1686	50.05	4.08	2813	607	317	0.6
Bentley F_n	3258	1670	48.95	4.44	2853	607	306	0.6
Punt & Smith	3296	1814	43.28	5.83	2790	607	308	0.8

Table 17b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.06	-0.04	0.35	-0.72	-0.21	-0.04	-0.30
Bentley's gradient	0.02	-0.17	-0.26	-0.25	-0.23	-0.02	-0.17
Haist	0.05	-0.36	-0.46	0.45	-0.70	-0.06	-0.25
Geromont & Glazer	-0.02	-0.21	-0.67	-0.22	-0.26	-0.01	-0.14
Bentley F_n	-0.02	-0.22	-0.63	-0.33	-0.28	-0.01	-0.17
Punt & Smith	-0.01	-0.15	-0.44	-0.75	-0.25	-0.01	-0.16

Table 17c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	0.01	-0.14	-3.98	-0.02	-0.38	0.00	-0.15
Cooke	-0.01	-0.10	-1.29	-0.03	-0.26	0.00	-0.17
Bentley's gradient	-0.01	-0.16	-1.38	-0.04	-0.27	0.00	-0.17
Haist	0.00	-0.25	-1.09	-0.08	-0.33	-0.01	-0.28
Geromont & Glazer	-0.01	-0.14	-1.21	-0.03	-0.27	0.00	-0.15
Bentley Fn	-0.01	-0.12	-1.05	-0.03	-0.26	0.00	-0.15
Punt & Smith	0.00	-0.13	-1.42	0.04	-0.31	0.00	-0.14

5.4.2.5 Episodic mortality

Table 18 summarises the results from trial r5, in which episodic mortality was simulated.

The crash rate was 6% for Haist's rule, 1.5% for Punt & Smith's rules, less than 1% for all others. For this situation, all indicators were worse than for the base case for all rules, except for the Haist rule's *Bav*, although the differences for *Bav* and *Cav* were small. Especially noticeable are the smaller *Bmin*, and larger *lowB* and *Bdiff* indicators, as one would expect with random mortalities. Most rules performed worse than the benchmark for most indicators.

Table 18a: Summary of indicators from 1000 runs in robustness trial r5.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3156	1496	29.87	4.98	2591	581	270	0.0
Cooke	3412	1317	34.75	8.13	3254	553	121	0.1
Bentley's gradient	3308	1191	49.18	5.44	3187	560	214	0.2
Haist	3508	1145	49.19	2.20	3806	529	285	6.4
Geromont & Glazer	3098	1084	60.07	5.08	3136	572	241	0.9
Bentley F_n	3098	1120	60.44	5.74	3154	571	221	0.6
Punt & Smith	3130	1127	57.42	7.01	3163	571	235	1.5

Table 18b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.08	-0.12	-0.16	-0.63	-0.26	-0.05	-0.55
Bentley's gradient	0.05	-0.20	-0.65	-0.09	-0.23	-0.04	-0.21
Haist	0.11	-0.23	-0.65	0.56	-0.47	-0.09	0.06
Geromont & Glazer	-0.02	-0.28	-1.01	-0.02	-0.21	-0.02	-0.11
Bentley F_n	-0.02	-0.25	-1.02	-0.15	-0.22	-0.02	-0.18
Punt & Smith	-0.01	-0.25	-0.92	-0.41	-0.22	-0.02	-0.13

Table 18c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	-0.05	-0.40	-3.96	-0.52	-0.60	-0.05	-0.38
Cooke	-0.04	-0.42	-3.07	-0.46	-0.52	-0.06	-0.61
Bentley's gradient	-0.04	-0.43	-2.09	-0.36	-0.47	-0.06	-0.42
Haist	0.00	-0.37	-1.35	-0.30	-0.33	-0.09	-0.26
Geromont & Glazer	-0.06	-0.45	-1.66	-0.29	-0.42	-0.06	-0.36
Bentley Fn	-0.06	-0.41	-1.53	-0.32	-0.40	-0.06	-0.39
Punt & Smith	-0.05	-0.46	-2.22	-0.16	-0.49	-0.06	-0.35

5.4.2.6 Increased random error

The first trial with increased stochastic error is summarised in Table 19. The benchmark crash rate stayed at zero, but that for the other rules ranged from 4% for Cooke's rule to 28% for Haist's rule.

Median Bav did not change much from the base case, but the *Bmin* and *lowB* indices were much worse. Median Cav was slightly lower and *Cmin* much lower (Table 19c). Most rules performed worse than the benchmark for all indicators, with some exceptions in the median and low biomass indices, and a better *AAV* from Haist's rule.

Table 19a: Summary of indicators from 1000 runs in robustness trial r6a.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3240	1700	30.86	6.61	3215	596	287	0.0
Cooke	3487	888	52.51	12.53	4425	562	20	3.6
Bentley's gradient	3373	573	92.89	8.40	4379	565	117	9.5
Haist	3589	159	189.55	3.60	5478	512	114	28.0
Geromont & Glazer	3154	485	101.17	8.41	4486	575	122	9.8
Bentley F_n	3136	433	110.67	9.37	4479	573	106	12.1
Punt & Smith	3345	959	66.44	8.51	4304	569	182	7.2

Table 19b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.08	-0.48	-0.70	-0.90	-0.38	-0.06	-0.93
Bentley's gradient	0.04	-0.66	-2.01	-0.27	-0.36	-0.05	-0.59
Haist	0.11	-0.91	-5.14	0.46	-0.70	-0.14	-0.60
Geromont & Glazer	-0.03	-0.71	-2.28	-0.27	-0.40	-0.03	-0.58
Bentley F_n	-0.03	-0.75	-2.59	-0.42	-0.39	-0.04	-0.63
Punt & Smith	0.03	-0.44	-1.15	-0.29	-0.34	-0.05	-0.37

Table 19c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	-0.02	-0.32	-4.12	-1.02	-0.99	-0.02	-0.34
Cooke	-0.02	-0.61	-5.16	-1.26	-1.07	-0.04	-0.94
Bentley's gradient	-0.02	-0.73	-4.84	-1.10	-1.03	-0.05	-0.68
Haist	0.03	-0.91	-8.04	-1.13	-0.92	-0.12	-0.70
Geromont & Glazer	-0.04	-0.75	-3.47	-1.13	-1.02	-0.05	-0.67
Bentley Fn	-0.05	-0.77	-3.62	-1.16	-0.98	-0.06	-0.70
Punt & Smith	0.01	-0.54	-2.72	-0.41	-1.02	-0.06	-0.49

When only the productivity process error was increased, and the two other errors left at the base case values (trial r6b, Table 20), the crash rate was 1–16%. The major changes from the base case were smaller *Bmin* and *Cmin* and larger *lowB* indices.

As in trial r6a, most rules performed worse than the benchmark for all indicators, with some exceptions in the median and low biomass indices.

Table 20a: Summary of indicators from 1000 runs in robustness trial r6b.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3248	1881	24.35	6.25	3033	600	335	0.0
Cooke	3516	1479	32.83	8.37	3747	568	157	1.2
Bentley's gradient	3423	1224	55.26	5.86	3883	574	226	2.6
Haist	3612	383	98.74	2.39	5020	533	226	16.2
Geromont & Glazer	3214	1109	68.08	5.61	3945	585	232	2.8
Bentley F_n	3208	1101	66.37	6.32	3947	586	221	3.3
Punt & Smith	3310	1355	48.43	6.85	3745	584	257	1.3

Table 20b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.08	-0.21	-0.35	-0.34	-0.24	-0.05	-0.53
Bentley's gradient	0.05	-0.35	-1.27	0.06	-0.28	-0.04	-0.33
Haist	0.11	-0.80	-3.05	0.62	-0.66	-0.11	-0.33
Geromont & Glazer	-0.01	-0.41	-1.80	0.10	-0.30	-0.03	-0.31
Bentley F_n	-0.01	-0.41	-1.73	-0.01	-0.30	-0.02	-0.34
Punt & Smith	0.02	-0.28	-0.99	-0.10	-0.23	-0.03	-0.23

Table 20c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	-0.02	-0.24	-3.04	-0.91	-0.88	-0.02	-0.23
Cooke	-0.02	-0.35	-2.85	-0.51	-0.75	-0.03	-0.49
Bentley's gradient	-0.01	-0.42	-2.47	-0.46	-0.80	-0.04	-0.39
Haist	0.03	-0.79	-3.71	-0.41	-0.76	-0.09	-0.41
Geromont & Glazer	-0.02	-0.43	-2.01	-0.42	-0.78	-0.04	-0.38
Bentley Fn	-0.02	-0.42	-1.77	-0.46	-0.75	-0.04	-0.39
Punt & Smith	0.00	-0.35	-1.71	-0.13	-0.76	-0.04	-0.29

When the implementation error for catch was increased (trial r6c, Table 21), crash rates remained low except for Haist's rule (7%). The main effects were decreased *Bmin* and *Cmin* and increased *Bdiff*. The median biomass and catch were nearly unchanged. Most rules were worse than the benchmark for most indicators.

Table 21a: Summary of indicators from 1000 runs in robustness trial r6c.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3310	2245	10.71	3.81	2006	607	360	0.0
Cooke	3546	1873	15.95	6.28	2660	584	220	0.0
Bentley's gradient	3443	1687	27.72	4.48	2697	590	282	0.1
Haist	3563	1276	41.21	1.87	3731	565	284	6.8
Geromont & Glazer	3256	1519	36.15	4.43	2789	600	280	0.3
Bentley F_n	3243	1495	37.23	4.88	2796	601	271	0.3
Punt & Smith	3283	1703	27.92	6.29	2593	601	288	0.2

Table 21b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.07	-0.17	-0.49	-0.65	-0.33	-0.04	-0.39
Bentley's gradient	0.04	-0.25	-1.59	-0.18	-0.34	-0.03	-0.22
Haist	0.08	-0.43	-2.85	0.51	-0.86	-0.07	-0.21
Geromont & Glazer	-0.02	-0.32	-2.38	-0.16	-0.39	-0.01	-0.22
Bentley F_n	-0.02	-0.33	-2.48	-0.28	-0.39	-0.01	-0.25
Punt & Smith	-0.01	-0.24	-1.61	-0.65	-0.29	-0.01	-0.20

Table 21c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	0.00	-0.10	-0.78	-0.16	-0.24	-0.01	-0.17
Cooke	-0.01	-0.18	-0.87	-0.13	-0.24	-0.01	-0.29
Bentley's gradient	0.00	-0.20	-0.74	-0.12	-0.25	-0.01	-0.23
Haist	0.02	-0.29	-0.97	-0.11	-0.30	-0.03	-0.26
Geromont & Glazer	-0.01	-0.23	-0.60	-0.12	-0.26	-0.01	-0.25
Bentley Fn	-0.01	-0.22	-0.56	-0.13	-0.24	-0.01	-0.25
Punt & Smith	-0.01	-0.18	-0.56	-0.04	-0.22	-0.01	-0.20

When the CPUE observation error was increased (trial r6d, Table 22), crash rates were low except for Haist's rule (4%). The main changes were in *Bmin, Cmin* and *AAV*. Median catch and biomass did not change much, and the *lowB* index remained low.

Table 22a: Summary of indicators from 1000 runs in robustness trial r6d.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3323	2495	5.69	3.26	1629	610	436	0.0
Cooke	3563	1955	13.36	9.30	2572	584	226	0.0
Bentley's gradient	3442	1866	23.29	6.54	2473	593	312	0.2
Haist	3590	1653	27.74	2.85	3265	568	326	3.7
Geromont & Glazer	3273	1698	29.84	6.87	2582	602	314	0.1
Bentley F_n	3268	1529	36.81	7.43	2827	601	276	0.0
Punt & Smith	3388	1932	19.03	8.08	2458	598	319	0.1

Table 22b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.07	-0.22	-1.35	-1.85	-0.58	-0.04	-0.48
Bentley's gradient	0.04	-0.25	-3.09	-1.01	-0.52	-0.03	-0.28
Haist	0.08	-0.34	-3.87	0.13	-1.00	-0.07	-0.25
Geromont & Glazer	-0.02	-0.32	-4.24	-1.11	-0.59	-0.01	-0.28
Bentley F_n	-0.02	-0.39	-5.46	-1.28	-0.74	-0.01	-0.37
Punt & Smith	0.02	-0.23	-2.34	-1.48	-0.51	-0.02	-0.27

Table 22c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	0.00	0.00	0.06	0.00	-0.01	0.00	0.00
Cooke	0.00	-0.14	-0.57	-0.67	-0.20	-0.01	-0.27
Bentley's gradient	0.00	-0.11	-0.46	-0.63	-0.14	-0.01	-0.15
Haist	0.03	-0.09	-0.32	-0.69	-0.14	-0.03	-0.15
Geromont & Glazer	0.00	-0.13	-0.32	-0.74	-0.17	-0.01	-0.16
Bentley Fn	-0.01	-0.20	-0.54	-0.72	-0.25	-0.01	-0.24
Punt & Smith	0.02	-0.07	-0.07	-0.34	-0.16	-0.01	-0.11

5.4.2.7 Increasing *q*

In this trial *q* increased during each run. The effects were large: except for *Cav*, all indicators were much worse than the base case for all rules. All rules performed considerably worse than the benchmark for *Bmin, AAV, Bdiff* and *Cmin*. Crash rates varied from 2% for Cooke's rule to 32% for Haist's rule (Table 23).

A typical set of results is shown in Figure 27.

Table 23a: Summary of indicators from 1000 runs in robustness trial r7.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin	crash rate (%)
Benchmark	2544	1359	88.18	4.64	2349	639	399	0.0
Cooke	2739	435	128.97	9.70	3628	618	0	2.0
Bentley's gradient	2644	433	149.33	5.89	3454	620	149	4.7
Haist	2734	207	226.71	2.48	4474	584	145	31.9
Geromont & Glazer	2467	382	169.00	5.64	3384	619	164	7.2
Bentley F_n	2478	372	172.96	6.33	3415	618	147	6.7
Punt & Smith	2596	739	127.13	6.35	3137	626	253	3.6

Table 23b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median Cav	20% Cmin
Cooke	0.08	-0.68	-0.46	-1.09	-0.54	-0.03	-1.00
Bentley's gradient	0.04	-0.68	-0.69	-0.27	-0.47	-0.03	-0.63
Haist	0.07	-0.85	-1.57	0.47	-0.90	-0.09	-0.64
Geromont & Glazer	-0.03	-0.72	-0.92	-0.22	-0.44	-0.03	-0.59
Bentley F_n	-0.03	-0.73	-0.96	-0.36	-0.45	-0.03	-0.63
Punt & Smith	0.02	-0.46	-0.44	-0.37	-0.34	-0.02	-0.37

Table 23c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	-0.23	-0.45	-13.63	-0.42	-0.45	0.05	-0.08
Cooke	-0.23	-0.81	-14.12	-0.75	-0.69	0.05	-1.00
Bentley's gradient	-0.23	-0.79	-8.39	-0.47	-0.60	0.04	-0.60
Haist	-0.22	-0.89	-9.82	-0.47	-0.56	0.00	-0.62
Geromont & Glazer	-0.25	-0.81	-6.47	-0.43	-0.53	0.02	-0.56
Bentley Fn	-0.25	-0.81	-6.23	-0.46	-0.51	0.02	-0.59
Punt & Smith	-0.22	-0.65	-6.12	-0.05	-0.47	0.03	-0.29

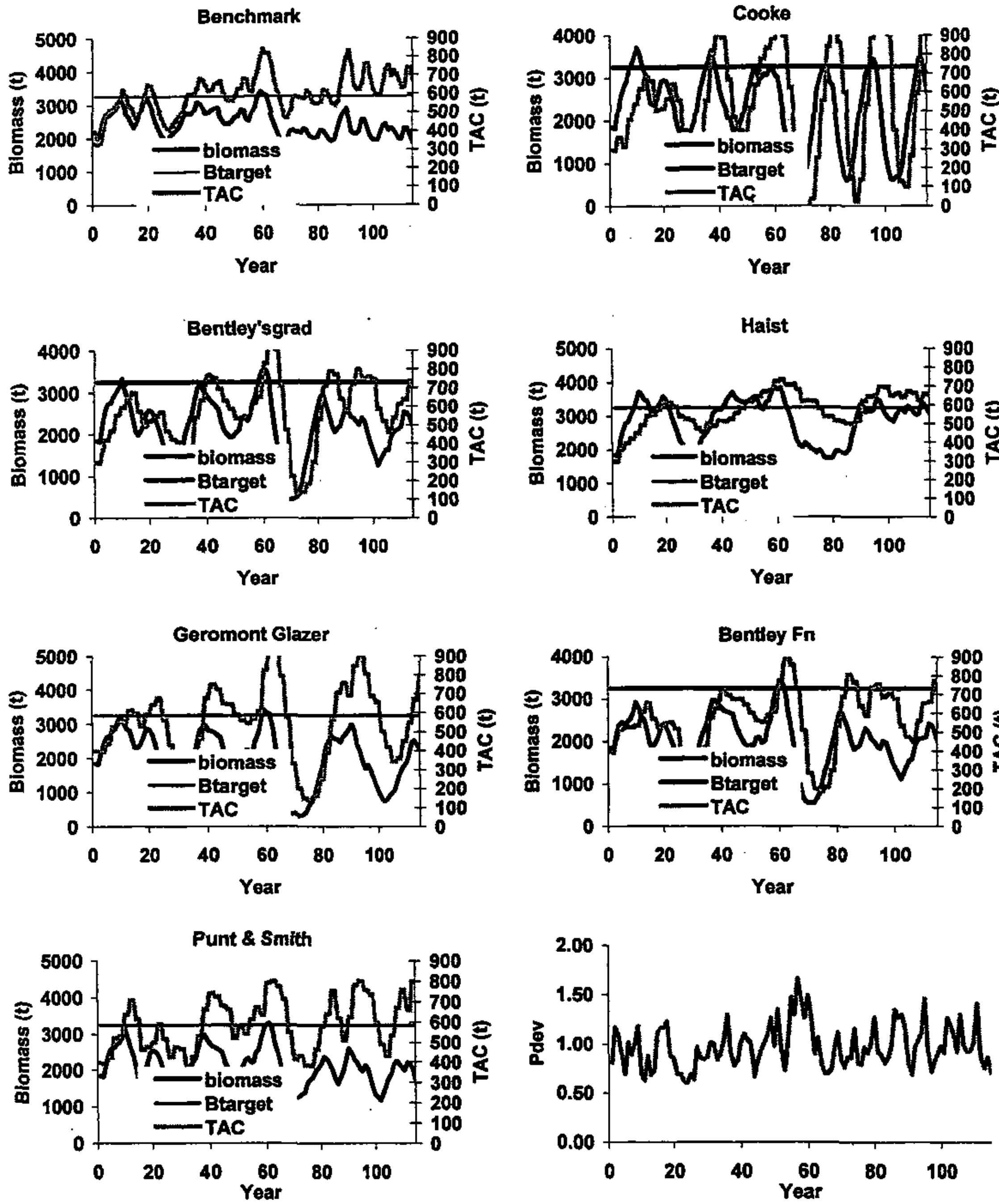

Figure 27: Typical runs from robustness trial r7, in which *q* increased during the run.

5.4.2.8 Unreported catches

Table 24 summarises the results from trial r8, in which a random unreported catch was simulated.

Haist's rule had a high crash rate, 18%, while all other rules were less than 1%. For all rules, *Bmin* and *lowB* were substantially worse than the base case; other indicators were usually slightly worse than the base case. Most rules were worse than the benchmark for most indicators except *Bav*, and except that Cooke's and Bentley's gradient rules were much better than the benchmark in *lowB*.

Table 24a: Summary of indicators from 1000 runs in robustness trial r8.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3073	2170	22.00	3.63	1660	621	430	0.0
Cooke	3378	1936	20.64	6.17	2381	599	272	0.2
Bentley's gradient	3321	1892	28.33	4.20	2261	603	358	0.6
Haist	3175	465	110.62	1.96	4050	589	233	18.3
Geromont & Glazer	3035	1536	52.91	4.31	2449	613	335	1.2
Bentley F_n	3030	1550	54.79	4.72	2434	613	329	0.6
Punt & Smith	3058	1715	43.99	6.38	2245	615	343	0.5

Table 24b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Cooke	0.10	-0.11	[redacted]	[redacted]	-0.43	-0.04	-0.37
Bentley's gradient	0.08	-0.13	[redacted]	-0.16	-0.36	-0.03	-0.17
Haist	0.03	[redacted]	[redacted]	0.46	[redacted]	-0.05	-0.46
Geromont & Glazer	-0.01	-0.29	-1.41	-0.19	-0.48	-0.01	-0.22
Bentley F_n	-0.01	-0.29	-1.49	-0.30	-0.47	-0.01	-0.24
Punt & Smith	0.00	-0.21	-1.00	[redacted]	-0.35	-0.01	-0.20

Table 24c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	-0.07	-0.13	-2.65	-0.11	-0.03	0.02	-0.01
Cooke	-0.05	-0.15	-1.42	-0.11	-0.11	0.02	-0.13
Bentley's gradient	-0.04	-0.10	-0.78	-0.05	-0.05	0.01	-0.03
Haist	-0.09	-0.74	-4.28	-0.16	-0.42	0.01	-0.39
Geromont & Glazer	-0.08	-0.22	-1.34	-0.09	-0.11	0.01	-0.10
Bentley Fn	-0.08	-0.19	-1.29	-0.09	-0.08	0.01	-0.09
Punt & Smith	-0.08	-0.18	-1.46	-0.06	-0.06	0.01	-0.04

5.4.2.9 Uncontrolled catches

In this trial we simulated four different situations. In the first situation, r9a, the uncontrolled catch was set in initial balance with the commercial catch, i.e., the combined catches were at about the level of population productivity, and there was no increase over time. Table 25 summarises the results. The crash rate of Haist rule is down to 0% from 2% in the base case (Table 25a). There was very little change between this result and the base case results (Table 25c). The largest change was seen for the *lowB* index for two rules, but this resulted from increases from small values in the base case.

Table 25a: Summary of indicators from 1000 runs in robustness trial r9a.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *C1av*	20% *C1min*	median *C2av*	20% *C2min*	median *Cav*	20% *Cmin*	crash rate (%)
Benchmark	3304	2531	5.44	3.07	1533	306	218	307	231	612	470	0.0
Cooke	3269	2286	12.76	5.67	1747	310	167	303	218	613	427	0.0
Bentley's gradient	3164	2124	21.45	3.77	1804	322	207	296	210	618	453	0.0
Haist	3256	2168	14.55	2.53	1992	305	194	303	212	608	463	0.0
Geromont & Glazer	3297	2341	10.48	3.52	1719	306	204	306	222	611	462	0.0
Bentley F_n	3295	2337	10.38	3.79	1738	306	198	305	222	611	457	0.0
Punt & Smith	3295	2368	9.72	6.28	1697	307	191	305	223	612	451	0.0%

Table 25b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *C1av*	20% *C1min*	median *C2av*	20% *C2min*	median *Cav*	20% *Cmin*
Cooke	-0.01	-0.10	-1.34	-0.85	-0.14	0.01	-0.23	-0.01	-0.06	0.00	-0.09
Bentley's gradient	-0.04	-0.16	-2.94	-0.23	-0.18	0.06	-0.05	-0.04	-0.09	0.01	-0.04
Haist	-0.01	-0.14	-1.67	0.18	-0.30	0.00	-0.11	-0.01	-0.08	-0.01	-0.01
Geromont & Glazer	0.00	-0.07	-0.93	-0.15	-0.12	0.00	-0.06	0.00	-0.04	0.00	-0.02
Bentley F_n	0.00	-0.08	-0.91	-0.24	-0.13	0.00	-0.09	-0.01	-0.04	0.00	-0.03
Punt & Smith	0.00	-0.06	-0.79	-1.04	-0.11	0.00	-0.12	-0.01	-0.03	0.00	-0.04

Table 25c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median *Cav*	20% *Cmin*
Benchmark	0.00	0.02	0.10	0.06	0.05	0.00	0.08
Cooke	-0.08	0.00	-0.50	-0.02	0.18	0.04	0.37
Bentley's gradient	-0.08	0.01	-0.35	0.06	0.17	0.03	0.23
Haist	-0.07	0.20	0.31	-0.50	0.30	0.04	0.21
Geromont & Glazer	0.00	0.19	0.54	0.11	0.22	0.01	0.24
Bentley Fn	0.00	0.23	0.57	0.12	0.23	0.01	0.27
Punt & Smith	0.00	0.13	0.46	-0.04	0.20	0.01	0.25

In the second situation, r9b, the uncontrolled catch was made larger, and may have been out of balance with the commercial catch, i.e., the combined catches may have been above the level of population productivity, but there was no increase over time. Results (Table 26) showed a very low percentage of crashes (0 to 1.1%), lower *Bav* and *Bmin* than in the base case, and much higher values for the *lowB* index. Average catch was slightly larger than in the base case, but the non-commercial catch was greater and the commercial catch less than in trial r9a.

Differences between the various rules and the base case were generally small, except that all rules had slightly better *lowB* indices, and Haist's rule had a lower commercial catch than the benchmark, and a lower total catch.

Table 26a: Summary of indicators from 1000 runs in robustness trial r9b.

	median *Bav*	20% *Bmin*	80% *lowB*	median *AAV*	80% *Bdiff*	median C1av	20% C1min	median C2av	20% C2min	median Cav	20% Cmin	crash rate (%)
Benchmark	2655	1726	64.44	3.98	1816	245	152	385	292	630	466	0.0
Cooke	2753	1681	63.44	7.24	1957	235	90	391	291	627	417	0.0
Bentley's gradient	2663	1494	82.51	4.19	2073	239	132	385	280	624	442	0.3
Haist	3004	1686	38.66	3.28	2346	200	108	410	294	610	457	1.1
Geromont & Glazer	2629	1496	85.99	3.97	2051	243	139	383	280	626	447	0.3
Bentley F_n	2632	1496	85.55	4.29	2064	243	135	383	280	626	447	0.4
Punt & Smith	2632	1548	82.65	6.60	1985	244	133	383	283	627	445	0.2

Table 26b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median	20%	80%	median	80%	median	20%	median	20%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*C1av*	*C1min*	*C2av*	*C2min*	*Cav*	*Cmin*
Cooke	0.04	-0.03	0.02	-0.82	-0.08	-0.04	-0.41	0.02	0.00	-0.01	-0.10
Bentley's gradient	0.00	-0.13	-0.28	-0.05	-0.14	-0.03	-0.13	0.00	-0.04	-0.01	-0.05
Haist	0.13	-0.02	0.40	0.17	-0.29	-0.18	-0.29	0.07	0.01	-0.03	-0.02
Geromont & Glazer	-0.01	-0.13	-0.33	0.00	-0.13	-0.01	-0.08	0.00	-0.04	-0.01	-0.04
Bentley F_n	-0.01	-0.13	-0.33	-0.08	-0.14	-0.01	-0.11	0.00	-0.04	-0.01	-0.04
Punt & Smith	-0.01	-0.10	-0.28	-0.66	-0.09	-0.01	-0.13	0.00	-0.03	0.00	-0.05

Table 26c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median	20%	80%	median	80%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*Cav*	*Cmin*
Benchmark	-0.20	-0.31	-9.69	-0.22	-0.12	0.03	0.07
Cooke	-0.23	-0.26	-6.44	-0.30	0.09	0.07	0.34
Bentley's gradient	-0.23	-0.29	-4.19	-0.05	0.04	0.04	0.20
Haist	-0.14	-0.07	-0.84	-0.95	0.18	0.04	0.19
Geromont & Glazer	-0.20	-0.24	-2.80	-0.01	0.07	0.03	0.20
Bentley Fn	-0.20	-0.22	-2.57	0.01	0.09	0.03	0.24
Punt & Smith	-0.20	-0.26	-3.63	-0.09	0.07	0.03	0.24

When the non-commercial catch allowed to increase (trial r9c; Table 27), crash rates increased to 3 – 34%. Cooke's and Haist's rules had much lower crash rates (3 – 5%) than the benchmark (22%). Biomass indicators were all lower than in trials 8a and 8b, and the *lowB* index was very high. Median total catch was near previous values, but the commercial catch was much decreased and the non-commercial catch much increased.

A notable result was that Haist's rule performed better than the benchmark rule in every indicator except commercial catch (it was the lowest of all rules). A typical set of results is shown in Figure 28.

Table 27a: Summary of indicators from 1000 runs in robustness trial r9c.

	median	20%	80%	median	80%	median	20%	median	20%	median	20%	crash
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*C1av*	*C1min*	*C2av*	*C2min*	*Cav*	*Cmin*	rate (%)
Benchmark	2108	245	232.63	5.44	3286	197	14	414	168	610	184	21.7
Cooke	2284	700	169.39	11.13	2906	180	0	448	271	627	351	2.7
Bentley's gradient	1992	215	280.20	5.24	3260	201	14	396	143	597	161	32.6
Haist	2341	699	151.50	3.24	2873	171	30	460	250	630	394	4.6
Geromont & Glazer	2044	214	267.89	4.76	3374	197	15	401	139	600	161	34.1
Bentley F_n	2054	214	264.65	5.13	3378	197	15	407	245	600	161	32.6
Punt & Smith	2065	218	258.09	6.79	3360	197	25	404	138	603	164	31.0

Table 27b: Showing the proportionate differences between indicators for the various rules indicated and the indicator from the benchmark rule (negative indicates worse, positive better).

	median	20%	80%	median	80%	median	20%	median	20%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*C1av*	*C1min*	*C2av*	*C2min*	*Cav*	*Cmin*
Cooke	0.08	1.85	0.27	-1.04	0.12	-0.09	-1.00	0.08	0.61	0.03	0.91
Bentley's gradient	-0.06	-0.12	-0.20	0.04	0.01	0.02	-0.05	-0.04	-0.15	-0.02	-0.12
Haist	0.11	1.85	0.35	0.41	0.13	-0.13	1.09	0.11	0.49	0.03	1.14
Geromont & Glazer	-0.03	-0.13	-0.15	0.12	-0.03	0.00	0.05	-0.03	-0.17	-0.02	-0.13
Bentley F_n	-0.03	-0.13	-0.14	0.06	-0.03	0.00	0.04	-0.02	0.46	-0.02	-0.13
Punt & Smith	-0.02	-0.11	-0.11	-0.25	-0.02	0.00	0.74	-0.02	-0.18	-0.01	-0.11

Table 27c: Showing the proportionate differences between indicators for the various rules indicated and the base case values for the same rule (negative indicates worse, positive better).

	median	20%	80%	median	80%	median	20%
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*Cav*	*Cmin*
Benchmark	-0.36	-0.90	-37.60	-0.66	-1.03	0.00	-0.58
Cooke	-0.36	-0.69	-18.86	-1.00	-0.36	0.07	0.13
Bentley's gradient	-0.42	-0.90	-16.62	-0.31	-0.51	0.00	-0.56
Haist	-0.33	-0.61	-6.23	-0.92	0.00	0.08	0.03
Geromont & Glazer	-0.38	-0.89	-10.84	-0.21	-0.52	-0.01	-0.57
Bentley Fn	-0.38	-0.89	-10.06	-0.18	-0.49	-0.01	-0.55
Punt & Smith	-0.38	-0.90	-13.45	-0.12	-0.58	-0.01	-0.55

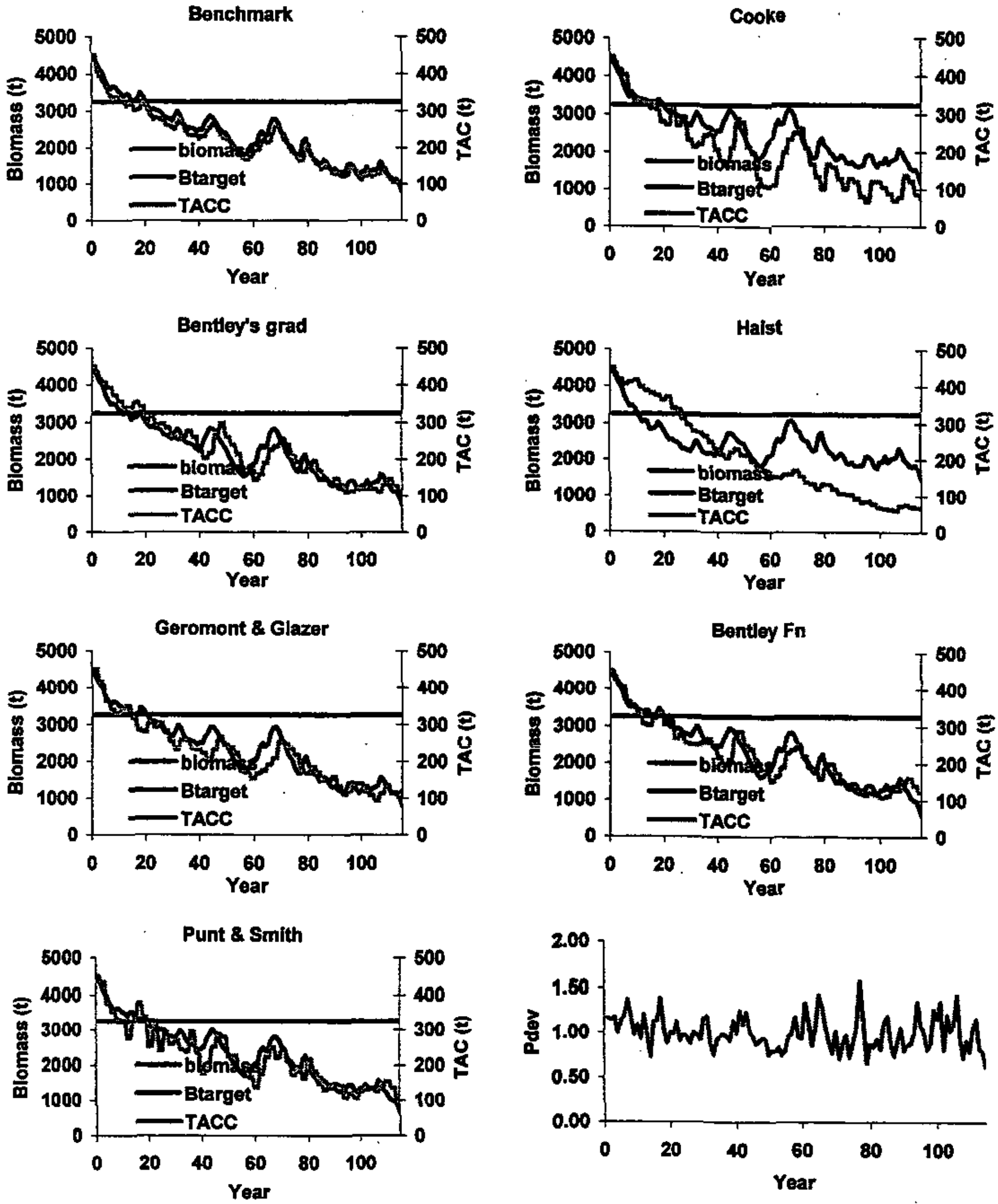

Figure 28: Typical runs from robustness trial r9c, in which *q* increased during the run.

When the non-commercial catch function from trial r9b was allowed to increase (trial r9d, Table 28), the crash rate increased to nearly 100% for all rules.

Table 28: Summary of indicators from 1000 runs in robustness trial r9d.

	median	20%	80%	median	80%	median	20%	median	20%	median	20%	crash
	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*C1av*	*C1min*	*C2av*	*C2min*	*Cav*	*Cmin*	rate (%)
Benchmark	897	133	673.20	9.65	2832	76	2	319	97	394	99	99.5
Cooke	1108	138	602.16	7.86	3052	64	0	373	103	441	103	97.7
Bentley's gradient	844	131	714.05	7.44	2864	71	2	309	95	380	98	99.5
Haist	1025	135	730.34	4.48	2974	64	1	345	94	413	101	98.0
Geromont & Glazer	841	131	712.52	6.54	2867	73	3	305	94	378	99	99.6
Bentley F_n	839	131	708.13	7.32	2878	74	3	494	312	380	98	99.6
Punt & Smith	850	132	700.72	3.59	2858	78	8	305	89	383	99	99.7

6. DISCUSSION

In the initial examination of harvest control rules we discovered that several rules do not behave well. The constant catch rule performed well only if the constant catch level was low. Whereas the well-behaved rules could produce a mean catch of more than 600 t in the base case simulations, the constant catch rule began to crash frequently at catches above 500 t (Figure 12). This was included only to demonstrate the trade-off between mean catch and catch stability.

The simple gradient-based rule of Geromont et al. (1999) for Namibian hake showed unacceptably high crash rates: 1% without a latent year and 15% with a latent year (see Table 5). The Fournier & Warburton (1988) rule showed either enormous periodic oscillations or unresponsiveness (see Figures 17 & 18) and could not be tuned to behave reasonably. The rule of Baldursson et al. (1996) also generated unstable cycles. The rule of Bentley et al. (unpublished results) demonstrated various poor behaviour types, depending on the parameters, and could not be tuned to perform well. Our own simple gradient-based rule behaved reasonably without a latent year, but became disconnected from biomass when a latent year was introduced.

The implementation lag and the latent year both contribute to unstable behaviour. Most rules produced better results, except for *AAV*, when implemented with no latent year. The effect of lag is suggested by the cycles in some rules (e.g., Cooke's rule in Figure 26), which may be caused by overshooting and undershooting the target because of the lag between signal and response.

These experiences underscore the need to explore a variety of potential harvest control rules before attempting to optimise a single version. Most studies do not do this. They also underscore the advantages of a simple system for running a number of rules and displaying the results. Although computational speed is poor, the Excel™ spreadsheet is ideal for comparing rules quickly (e.g., Figure 26).

It is, of course, unrealistic to model rules that are fixed in place for 100 years. A poorly performing rule would not be permitted to operate for long. However, it is important to evaluate rule behaviour over a long time frame to separate rules that do or do not have the potential to give stable performances.

Of the rules we chose to fine-tune and proceed further with, we included several types:

- a version of Cooke's (1999) rule for baleen whales, based on an arbitrary linear biomass-catch function,
- Bentley's gradient- and target-based rule,
- Haist's (2002) hoki rule using estimated surplus production and a target CPUE,
- Geromont & Glazer's ((1998) buffered Fn hake rule,
- Bentley's buffered Fn rule, and

- Punt & Smith's (1999) buffered Fn rule for gemfish.

Thus three of the rules chosen for further testing were Fn rules and the others were simple or hybrid rules. In the base case simulations (see Table 12), differences among these rules, and between the rules and the benchmark, were generally minor. (An exception was the small positive crash rate in Haist's rule, while all the other rules in this phase were crash-free.) If the base case simulations were reliable as representing reality, then any of these rules could be used in a management procedure for lobsters.

However, results from some of the robustness trials varied widely from the base case results. This testing confirmed the suggestion that robustness to structural uncertainties is more important than optimisation.

The most dramatic failures in the robustness trials occurred when a substantial uncontrolled catch was allowed to increase. In trial r9c, when an uncontrolled catch originally in balance with the controlled catch and with population productivity (trial r9a) was allowed to increase at a small rate (1% annually), biomass was badly depressed by the end of the runs for all rules including the benchmark (Figure 28), the crash rate increased to 3 – 34% (see Table 27a) and the average commercial catch declined from over 300 t (see Table 26a) to less than 200 t. When the original uncontrolled catch was already too high, as in trial r9b, allowing a small annual increase caused the model run to crash for nearly every run for every rule.

Interestingly, the Haist rule, which did not perform as well as some of the rules in many trials, showed the lowest crash rate in trial r9c. This illustrates that an overall best rule is very difficult to find except by specifying a narrow range of possible states of nature. If catches were uncontrolled in reality, this would be a good rule; if this robustness trial does not mimic a likely real situation, the rule is not a good one.

Some implementation error can be absorbed by the harvest control rules tested. For instance, even the base case has some implementation error in catches, but the rules perform satisfactorily (see Table 12). When this error was increased in trial r6c (see Table 21), mean biomass and catch indicators were nearly unchanged, although minimum biomass indicators deteriorated somewhat and the crash rate for one rule increased to 7%. Despite this tolerance, the results from trial r9 highlight the need for a management procedure to influence total catch effectively. Results were satisfactory only when the uncontrolled catch was both static and in balance with productivity. When the uncontrolled catch was too high for the productivity, or was allowed to increase, controlled catch was decreased by all rules and biomass also decreased, with population crashes when uncontrolled catch became high enough.

This result should not be surprising: the wide experience of fisheries management is that fisheries with increasing catches or effort tend to become over-fished and often seriously depleted unless catch or effort is controlled. The situation modelled here was not extreme: current estimates place the sum of recreational, traditional and illegal catches in the NSN stock at roughly the same level as those from the commercial fishery; they are not controlled and are thought to be increasing.

In such a situation, management procedures can be successful only if the management strategy can be applied to the whole catch.

Other robustness trials modelled effects that were much more benign. Trial r1 suggested that errors in the estimates of productivity parameters used to tune harvest control rules increase the variability of results, and decrease low and minimum biomass indicators, but do not generally compromise the behaviour of the rules tested. The crash rate increased to 6% (see Table 14) for one rule only, but other rules behaved acceptably, although worse than in the base case. Introducing a lag in productivity had little effect (see Table 15). Where *K* varied systematically, the Haist rule behaviour became unacceptable (see Table 16), but all other rules behaved acceptably. A regime shift caused the Haist rule's crash rate to increase to 6% (see Table 17), but otherwise all rules performed acceptably; similarly for episodic mortality (see Table 18). Unreported catches (see Table 24) caused high crash rates in one rule (see Table 24a) but have relatively minor effects on other rules.

If production variability were greater than in the base case simulations, performance is degraded and the crash rate increases (trial r6, Tables 19 and 20). Increased observation error and implementation error have much smaller effects (Tables 21 and 22). The implication is that, for obtaining robust results from simulation studies, production error should err on the side of over-estimation.

Increasing *q* had serious effects (see Table 23). This is a realistic trial, because increasing technology could easily cause inflation of CPUE over time. However, the trial is unrealistic in modelling the situation developing steadily over 100 years – a regular stock assessment using other data should be able to identify this problem and allow correction of the management procedure.

Among the harvest control rules, it is difficult to see much consistent difference. The exception to this is the Haist rule: it often showed a crash rate several times that of the others (see Tables 18 – 24) and in two trials was unacceptably high when the others were relatively low (see Tables 23 and 24). In fairness, in trial r9c (see Table 27) the Haist rule had a low crash rate while most others were high, including the benchmark rule. This behaviour mandates either removing the Haist harvest control rule as a candidate for use in a lobster management procedure, or further tuning this rule to try to eliminate crash rate problems.

Of the remaining rules, is any one rule a better performer than the rest? We ranked the performance of each rule for each indicator in the base case and all robustness trials, 1 for best and 5 for worst, giving the mean value to both rules in cases of ties. For the average biomass indicator, the value above the target biomass but closest to the target scored 1, the next highest (if any scored 2); after the values above the target were gone, the highest value below the target took the next score, etc. The *%Crash* indicator was not included, because a manager would probably want to reject rules above some threshold risk of crashing. The sums of scores are shown in Table 29.

Table 29: Scores for different rules for each of the indicators. For method see text.

	Bav	*Bmin*	*lowB*	*AAV*	*Bdiff*	*Cav*	*Cmin*	*C1av*	*C1min*	*C2av*	*C2min*	%Crash	*sum*
Benchmark	-	-	-	-	-	-	-	-	-	-	-	7.6	-
Cooke	54	23	23	90	48	67	85	17	24	10	9	6.7	450
Bentley's gradient	51	55	54	47	49	62	59	10	14	19	17	9.5	437
Geromont & Glazer	45	66	75	37	54	42	42	13	8	16	19	10.1	417
Bentley Fn	47	79	75	64	67	42	67	12	12	13	11	10.0	489
Punt & Smith	43	38	41	81	28	36	41	8	11	16	16	9.3	359

Overall, the Punt & Smith rule scored highest, although it did not have the lowest crash rate. This follows the prediction of Polacheck et al. (1999) that constant catch rate control rules work best.

Further work on management procedures for the rock lobster fishery will require a management approach that addresses the large uncontrolled catches.

7. ACKNOWLEDGMENTS

This work was conducted under Ministry of Fisheries contract CRA 2000-01. Thanks to Kevin Stokes and Vivian Haist for interesting discussions, and to David Gilbert for a very helpful and thorough review.

8. REFERENCES

Baldursson, F.M.; A. Danielsson & G. Stefansson. (1996). On the rational utilization of the Icelandic cod stock. *ICES Journal of Marine Science 53:* 643 – 658.

Bentley, N.; Breen, P.A.; Starr, P.J.; Sykes, D.R. (2002). Development and evaluation of decision rules for management of New Zealand rock lobster fisheries. *New Zealand Fisheries Assessment Report 2002/.*

Bentley, N.; Breen, P.A.; Starr, P.J. submitted. Design and evaluation of a revised management decision rule for red rock lobster fisheries (*Jasus edwardsii*) in CRA 7 and CRA 8. *New Zealand Fisheries Assessment Report 2002/xx.*

Breen, P. A.; Hilborn, R.; Maunder, M. N.; Kim, S. W. submitted. Effects of alternative control rules on the conflict between a fishery and a threatened sea lion. *Canadian Journal of Fisheries and Aquatic Sciences.*

Breen, P.A.; Starr, P.J.; Bentley, N. (2001). Rock lobster stock assessment for the NSN substock and the combined CRA 4 and CRA 5 areas. *New Zealand Fisheries Assessment Report 2001/7.* 42 pp.

Butterworth, D.S.; Cochrane, K.L.; De Oliveira, J.A.A. (1997). Management procedures: A better way to manage fisheries? The south African experience. pp. 83 – 90 *In* E.K. Pikitch, D.D. Huppert & M.P. Sissenwine (Eds.) Global trends: Fisheries management. *American Fisheries Society Symposium 20.*

Butterworth, D.S.; Punt, A.E. (1999). Experiences in the evaluation and implementation of management procedures. *ICES Journal of Marine Science 56:* 985 – 998.

Cochrane, K.L.; Butterworth, D.S.; De Oliveira, J.A.A.; Roel, B.A. (1998). Management procedures in a fishery based on highly variable stocks and with conflicting objectives: Experience in the South African pelagic fishery. *Reviews in Fish Biology and Fisheries 8:* 177 – 214.

Cooke, J.G. (1999). Improvement of fishery-management advice through simulation testing of harvest algorithms. *ICES Journal of Marine Science 56:* 797 – 810.

De Oliveira, J.A.A.; Butterworth, D.S.; Johnston, S.J. (1998). Progress and problems in the application of management procedures to South Africa's major fisheries. pp. 513 – 530 *In* T.J. Quinn, F. Funk, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan & C.-I. Zhang (Eds.) Fisheries Stock Assessment Models. Alaska Sea Grant College Program, AK-SG-98-01. University of Alaska, Fairbanks.

Fournier, D.A.; Warburton, A.R. (1988). Evaluating fisheries management models by simulated adaptive control - introducing the composite model. *Canadian Journal of Fisheries and Aquatic Sciences 46:* 1002 – 1012.

Francis, R.I.C.C.; Hurst, R.J.; Renwick, J.A. (2001). An evaluation of catchability assumptions in New Zealand stock assessments. *New Zealand Fisheries Assessment Report 2001/1.* 37 p.

Geromont, H.F.; De Oliveira, J.A.A.; Johnston, S.J.; Cunningham, C.L. (1999). Development and application of management procedures for fisheries in southern Africa. *ICES Journal of Marine Science 56:* 952 – 966.

Geromont, H.F.; Glazer, J.P. (1998). The 1998 revised management procedures for South African west coast hake resource. SFRI Document WG/11/98/D:H:49. 27 pp. (cited by Geromont et al. 1999).

Haist, V.R. (2002). Evaluation of a management procedure for New Zealand hoki fisheries. PhD thesis, University of Washington, Seattle USA.

Hilborn, R. (1979). Comparison of fisheries control systems that utilize catch and effort data. *Journal of the Fisheries Research Board of Canada 36:* 1477 – 1489.

Kirkwood, G.P. (1997). The revised management procedure of the International Whaling Commission. pp. 91-99 *In* Pikitch, E.K.; Huppert, D.D.; Sissenwine, M.P. (Eds.) Global trends: Fisheries management. *American Fisheries Society Symposium 20.*

McAllister, M.K.; Starr, P.J.; Restrepo, V.R.; Kirkwood, G.P. (1999). Formulating quantitative methods to evaluate fishery-management systems: What fishery processes should be modelled and what trade-offs should be made? *ICES Journal of Marine Science 56:* 900 – 916.

Patterson, K.; Cook, R.; Darby, C.; Gavaris, S.; Kell, L.; Lewy, P.; Mesnil, B.; Punt, A.; Restrepo, V.; Skagen, D.; Stefansson, G. (2001). Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries 2:* 125 – 157.

Polacheck, T.; Klaer, N.L.; Millar, C.; Preece, A.L. (1999). An initial evaluation of management strategies for the southern bluefin tuna fishery. *ICES Journal of Marine Science 56:* 811 – 826.

Punt, A.E. (1993). The comparative performance of production-model and *ad hoc* tuned VPA based feedback-control management procedures for the stock of Cape hake off the west coast of South Africa. pp. 283-299 *In* S.J. Smith, J.J. Hunt & D. Rivard (Eds.) Risk evaluation and biological reference points for fisheries management. *Canadian Special Publication of Fisheries and Aquatic Sciences 120.*

Punt, A.E.; Smith, D.M. (1999). Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). *ICES Journal of Marine Science 56:* 860-875

Robb, C.A.; Peterman, R.M. (1998). Application of Bayesian decision analysis to management of a sockeye salmon (*Oncorhynchus nerka*) fishery. *Canadian Journal of Fisheries and Aquatic Sciences 55:* 86 – 98.

Starr, P.J.; Breen, P.A.; Hilborn, R.; Kendrick, T.H. (1997). Evaluation of a management decision rule for a New Zealand rock lobster substock. *Marine and Freshwater Research 48(8):* 1093 – 1101.