

## BEYOND MSE: OPPORTUNITIES IN THE APPLICATION OF ATLANTIC BLUEFIN TUNA OPERATING MODELS

Tom Carruthers<sup>1</sup> and Laurence Kell<sup>2</sup>

### SUMMARY

*Management strategy evaluation (MSE) requires the description of plausible hypotheses for population and fishery dynamics, also known as operating models. When performance metrics are available, fishery management procedures may be identified that are robust to uncertainties. However when operating models and performance metrics can be established, they offer many additional opportunities beyond MSE and the selection of MPs, and can be used to quantify value of information, test experiment designs, identify appropriate management reference points and investigate enforcement strategies. We discuss some of the opportunities and limitations of operational modelling.*

### RÉSUMÉ

*L'évaluation de la stratégie de gestion (MSE) nécessite la description d'hypothèses plausibles pour la dynamique des populations et des pêcheries, également appelées modèles opérationnels. Lorsque des mesures de la performance sont disponibles, on peut identifier des procédures de gestion des pêcheries qui soient solides face aux incertitudes. Cependant, lorsque des modèles opérationnels et des mesures de la performance peuvent être établis, ils offrent de nombreuses opportunités supplémentaires au-delà de la MSE et de la sélection des procédures de gestion et ils peuvent servir à quantifier la valeur de l'information, tester des modèles expérimentaux, identifier des points de référence de gestion appropriés et étudier les stratégies d'application. Nous discutons de certaines des opportunités et des limites de la modélisation opérationnelle.*

### RESUMEN

*La evaluación de estrategias de ordenación (MSE) requiere la descripción de hipótesis plausibles para la dinámica de la pesquería y de la población, también conocidas como modelos operativos. Cuando se dispone de mediciones del rendimiento, pueden identificarse procedimientos de ordenación pesquera que sean robustos ante las incertidumbres. Sin embargo, cuando pueden establecerse modelos operativos y mediciones del rendimiento, ofrecen muchas oportunidades adicionales más allá de la MSE y la selección de procedimientos de ordenación (MP), y pueden utilizarse para cuantificar el valor de la información, probar diseños experimentales, identificar puntos de referencia de ordenación adecuados e investigar estrategias de ejecución. Se discuten algunas de las oportunidades y limitaciones de los modelos operativos.*

### KEYWORDS

*Population modelling, spatial analysis, data collections, age composition, aerial surveys, catch statistics, fishery statistics, fishing effort, size composition, tagging*

<sup>1</sup> IOF, 2202 Main Mall, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z4. t.carruthers@fisheries.ubc.ca

<sup>2</sup> International Commission for the Conservation of Atlantic Tunas, Calle Corazón de María, 8, 28002 Madrid, Spain

## 1 Introduction

A Management Strategy Evaluation (MSE, Butterworth 1999, Cochrane 1998,) approach has been proposed for Atlantic bluefin tuna (SCRS 2013) as a suitable framework for providing robust management advice consistent with the precautionary approach (GBYP 2014). A principal task in the construction of an MSE framework is the development of operating models which represent credible hypotheses for population and fishery dynamics. Operating models are typically fishery stock assessment models which are fitted to data to ensure that model assumptions and estimated parameters are empirically credible (Punt *et al.* 2014, e.g. CCSBT 2011).

At the heart of an MSE analysis is a closed-loop simulation, so called because the advice arising from management procedures (e.g. a TAC) feeds back into the known, simulated population dynamics. This approach has been applied widely to reveal trade-offs in performance and identify management procedures (MPs) (Punt *et al.* 2014, Carruthers *et al.* 2015) that are robust to uncertainty in observations and fishery dynamics. A wide range of management procedures can be tested that include assessments linked to harvest control rules, spatio-temporal closures, size limits and allocation schemes.

However, beyond the core MSE there are many other valuable uses for an established set of operating models and performance metrics (quantifiable definition of what is desirable and undesirable in the context of fisheries management). In general it is not possible to establish the bias of experiments or analyses without perfect information of the relevant quantities (e.g. simulated depletion, stock abundance, stock trajectory, fishing mortality rate etc.). It follows that simulation offers a foundation for evaluating and optimizing many components of fishery management system such as data collection, enforcement and reference points.

## 2 Value of information analysis

Arguably one of the most important uses of operating models is quantifying the value of various sources of information. There are three distinct types of value of information analysis: the value of better data, the value of additional data and the cost of current uncertainties.

### *Value of better data*

The value of better data considers more precise and less biased data arising from better experimental design, increased sampling intensity or more sophisticated data filtering and processing techniques. Analysis of better data can be achieved by quantifying the marginal benefit of increasing the precision or accuracy of the various data types (e.g. a relative abundance index, annual catches, catch composition) and establishing those that most strongly affect the yields (or any metric of utility) derived from a particular management system (e.g. **Figure 1**).

### *Value of additional data*

Value of additional data examines the potential benefit of other management procedures that are possible given the collection of new types of data, for example, close-kin genetic tagging (Bravington *et al.* 2013), gene tagging, hydro-acoustic detection (Goñi *et al.* 2016, Canals *et al.* 2016) larval surveys (Ingram *et al.* 2015) or additional aerial surveys (Bonhommeau 2010). If defensible observation models can be designed for new types of data, the efficacy of alternative management procedures that use these data can be tested to establish whether they are worthy of collection and processing (e.g. **Figure 2**).

### *Cost of current uncertainties*

Post-hoc analysis of MSE results often reveals gradients in performance with respect to particular parameters of the operating model; most often: natural mortality rate, steepness of the stock recruitment curve, stock depletion, the size selectivity of fishing, annual increases in fishing efficiency and spatial targeting. This analysis doesn't identify specific data to be collected or improved but simply highlights where operating model uncertainty may lead to selection of MPs that are worse than other MPs over sub-ranges of model parameters (e.g. **Figure 3**).

This problem can be rephrased as the yield lost due to not using an MP that may perform better but cannot be selected due to risks associated with parameter uncertainty. For example the DCAC MP may provide higher expected yields over an MSE projection than a delay-difference (DD) MP but only if it is certain that stock depletion is above 20% of unfished (**Figure 3**). The cost of current uncertainties is driven by asymmetry in performance among MPs; for example a more conservative MP may often be selected in the presence of higher uncertainty (consistent with the precautionary principle).

### **3 Design of stock assessments**

There are a range of protocols for weighing the various data used in a stock assessment (e.g. relative abundance indices, size composition data, catch observations, tagging data etc.), including iterative reweighting (McAllister and Ianelli 1997) and prioritization of fits to certain data (Francis 2011). However the existing data weighting approaches are ad-hoc and are not designed to establish robust management outcomes over a projected period. For example, over a 50 year projection it may be more important to robustly fit an index of relative abundance than precisely estimate size selectivity of fishing (the hypothesis of Francis 2011). Closed - loop simulation offers the most theoretically consistent basis for testing a range of data-weighting schemes and selecting one based on established management performance metrics. The operating models for Atlantic bluefin tuna can be used in a peer-reviewed investigation of the success of various proposed data weighting schemes for various stock assessments.

Similarly to data-weighting decisions about the appropriate complexity of stock assessment models are generally based on information theory and the notion that the model should provide a parsimonious fit to the observed data (e.g. AIC, BIC and other diagnostics of model fit). However these model selection criteria are entirely unrelated to the long-term interests of fisheries managers. A consistently biased model that fits the data poorly may never chronically overfish, leading to healthy stock sizes, higher catch rates and larger fish caught (may satisfy a number of possible management performance measures). There is no coherent reason to assume that the statistical properties of model fit are related in any way to the emergent management performance of a particular stock assessment. Perhaps catches are chronically misreported or relative abundance indices are hyperstable, in which case a model that appears in a single year to be statistically suspect may provide appropriate management advice. Using operating models to weight data is also considerate of the unique conditions of each assessment setting. Take the example of a relative abundance index that could be strongly hyperstable; when does the recommendation of Francis (2011) to prioritize fit to the index hold true and when should it be ignored? The unpalatable truth may be that data weighting is context specific and there are few general rules.

Operational modelling provides a meaningful opportunity to move away from the current, possibly myopic statistical approach to model selection, perhaps leading to the use of simpler, more tractable and easier to understand stock assessments that can meet management performance requirements robustly.

### **4 Optimizing for spatial fleet structure and allocation**

There is a diverse range of fishing métiers operating on the tunas of the Atlantic (Mediterranean trap, Canadian rod and reel etc.). Fleet heterogeneity arises not only from varying gear configuration but also the magnitude and spatio-temporal arrangement of effort. Rather than testing prescribed spatio-temporal controls and allocation schemes it may be possible to directly optimize for these under the various operating model scenarios, in such a way that yields may be improved whilst providing protection for less resilient stocks.

### **5 Evaluating methods of data processing**

Catch rate data provide the longest time series of relative abundance information for most pelagic tunas including bluefin. It is common for these CPUE data to be standardized whereby a linear model is used to remove the various confounding factors that affect CPUE other than population density (e.g. bait type, season, depth etc). Similarly to the selection of assessment models, standardization models are generally chosen that satisfy model selection criteria such as AIC. However simulation testing reveals that selecting standardization models this way is flawed and often leads to the selection of models that produce spurious estimates of relative abundance (Carruthers et al. 2010). Operating models for bluefin tuna may be used to establish robust CPUE standardization models and identify the potential for non-linearity in the relationship between stock size and the index.

For certain data sources (e.g. total catch-at-age data), the majority of data points are manually imputed using ad-hoc rules (for example uprating to total catches assuming age composition of similar fleets operating in similar times and areas). These approaches can be time consuming and opaque but also hard to validate. Operating models may be used to establish automated data-imputation algorithms that are robust and can be applied rapidly (e.g. Carruthers and Kell 2016).

## 6 Determining appropriate fisheries management reference points and fishery certification standards

Fishery reference points (e.g. biomass targets and limits) are often established without consideration of what is meaningful and achievable given the case-specific fishery and population dynamics (Hilborn 2002). For example in some management settings fisheries are considered healthy only when spawning biomass is above MSY levels. For short-lived species with highly variable recruitment, there is commensurate variability in unfished biomass which masks the management performance when this is expressed in terms of biomass. In such cases reference points relating to exploitation rate may be more appropriate. Operating models may be used to test the appropriateness of candidate management reference points by quantifying the success of MPs (control rules) that use these to make policy recommendations. Simple tests include evaluating whether management performance limits (e.g. a 50% probability of a stock above 50% biomass at MSY) and targets can be met given fishing at fractions of a known simulated level, for example no fishing, 50%  $F_{MSY}$  fishing and  $F_{MSY}$  fishing.

Eco-labelling and seafood certification is rapidly expanding and in many instances has improved the profitability of fishing operations whilst providing guidance to ensure sustainability, albeit with uncertain environmental benefits (MRAG 2009). The fishery certification standards (for example SG100, SG80 and SG60 of the Marine Stewardship Council) have previously used the outputs of stock assessments to inform certification purposes. However a new initiative being investigated for less data-rich fisheries is whether certification standards can be met by management procedures that do not provide explicit estimates of stock status and trajectory (Carruthers and Agnew 2016). Under this new certification paradigm, operating models may be used to certify fisheries that do not have an assessment, for example by quantifying the likelihood of meeting a certification standard using a size limit. It follows that operating models may offer an alternative approach to meeting certification standards in situations where a single stock assessment is not considered representative of the full range of uncertainty.

## 7 Testing enforcement strategies

In the multi-stock Atlantic bluefin tuna fishery there is potential for high spatio-temporal heterogeneity in fishing mortality rates (e.g. fishing on spawning grounds). This may be a particularly issue for certain fishing métiers operating on less resilient stocks. Operating models offer a basis for testing enforcement strategies. Given the variable costs of enforcement (Davis *et al.* 2015), where and when should monitoring be prioritized and for what métiers (e.g. Bastardie *et al.* 2014)? Enforcement strategies may be linked to value of information analysis to target the most critical data reporting issues.

## 8 Discussion

In this paper we focus primarily on the potential opportunities arising from establishing operating models for Atlantic bluefin tuna. Here we discuss some of the potential problems and limitations.

Arguably the most important limitation is the likely failure to generate data as irregular as real fishery data. In many cases what is considered observation error encompasses fundamental model misspecification that is not well simulated by simplistic observation error models. For example, certain fleets such as purse seiners operate on aggregations of bluefin tuna that are often size-monospecific (shoaling fish of comparable size) leading to size observations that are strongly non-independent. Other fleets operate on more mixed size classes (e.g. US rod and reel) but this may change to homogeneous size classes during certain times of year (e.g. spawning). Fisheries may high-grade fish depending on temporally varying market conditions and catch rates, retaining the largest caught without reporting the mortality of smaller fish and there may be complex spatial targeting for fleets operating on multiple species such as pelagic longliners. Fishing dynamics such as these produce irregular data sets that may be poorly recreated by naïve observation error models (e.g. generating size composition from multivariate logistic or multinomial models). The result could be misleading value of information analyses and model selection due to overstating the information content of the various data.

Simulation testing is generally only appropriate for rapid, robust, automated and reproducible methods. This precludes the simulation evaluation of subjective approaches such as productivity-susceptibility analysis (PSA, Patrick *et al.* 2009) and to a certain extent a realistic stock assessment process which includes many opportunities for experts to intervene where problems are apparent. It follows that simulation may provide an overly pessimistic view of current practices that rely on subjective judgement. Depending on how computationally intensive a method is, simulation evaluation may not be tractable. However this limitation raises the interesting question of whether non-objective, non-reproducible approaches should be used if it is not

possible to evaluate their biases and more generally their performance. If a conventional stock assessment process is not objective by definition and cannot be automated, can it be considered best available scientific practice?

The principal driver behind developing operating models for Atlantic bluefin tuna is the identification of management procedures that are robust to the considerable uncertainties in population and fishing dynamics (Anon. 2014). More broadly, operational modelling provides a coherent approach for improving the efficiency of fisheries management by prioritizing the most valuable science and identifying the most appropriate methods.

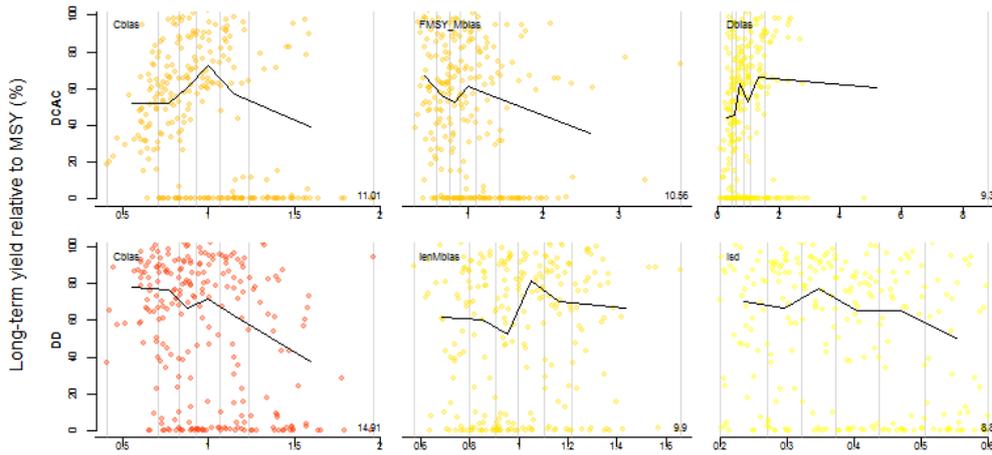
### **Acknowledgements**

This work was carried out by TC under the provision of the ICCAT Atlantic Wide Research Programme for Bluefin Tuna (GBYP), funded by the European Union, several ICCAT CPCs, the ICCAT Secretariat and by other entities (see: <http://www.iccat.int/GBYP/en/Budget.htm>). The contents of this paper do not necessarily reflect the point of view of ICCAT or other funders and in no ways anticipate ICCAT future policy in this area.

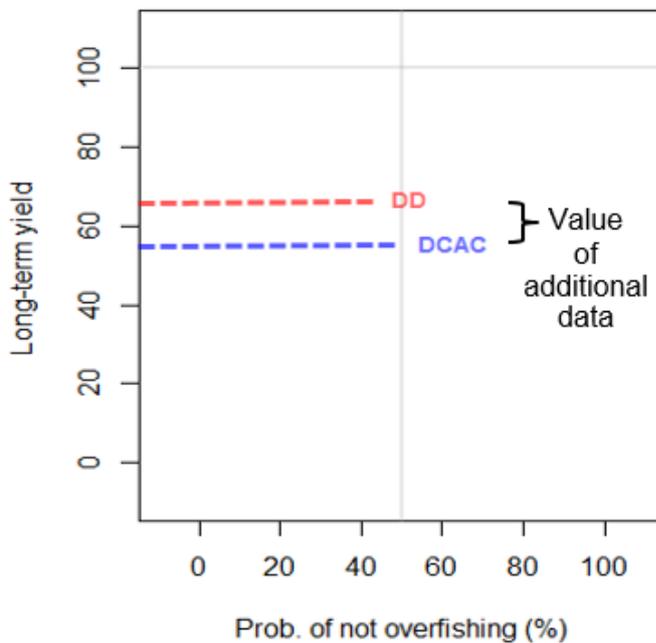
## References

- Anon. 2014. Report of the 2013 bluefin meeting on biological parameters review (Tenerife, Spain - May 7 to 13, 2013).
- Bastardie, F., Nielsen, J.R., Meithe, T. 2014. DISPLACE: a dynamic, individual-based model for spatial planning and effort displacement - integrating underlying fish population dynamics. *Can. J. Fish. Aquat. Sci.* 71: 366-386.
- Bonhommeau, S., Farrugio, H., Poisson, F, Fromentin, J-M. 2010. Aerial surveys of bluefin tuna in the western Mediterranean Sea: retrospective, prospective, perspective. *Col. Vol. Sci. Pap. ICCAT*, 65(3): 801-811.
- Bravington, M. V., Grewe, P.G., Davies, C.R. 2013. Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers. FRDC Report 2007/034. CSIRO, Australia.
- Butterworth, D.S., Punt, A.E., 1999. Experiences in the evaluation and implementation of management procedures. *ICES J. Mar. Sci.* 56, 985-998.
- Canals M., Balguerías E., Stokesbury M., Whoriskey F., Sánchez A., Medina A., Abascal F.J. and Aranda G. 2016. An acoustic telemetry curtain across the Strait of Gibraltar? SCRS/2015/056.
- Carruthers, T.R. and Agnew, D. 2016. Using simulation to determine standard requirements for recovery rates of fish stocks. *Marine Policy.* 73, 146-153.
- Carruthers, T.R. and Kell, L. 2016. Quantifying uncertainty due to data processing in age-structured stock assessments. Paper for submission to *ICES Journal of Marine Science*.
- Carruthers, T.R., Ahrens, R. and McAllister, M. 2010. Simulating spatial dynamics to evaluate methods of deriving relative abundance indices for tropical tunas. *Canadian Journal of Fisheries and Aquatic Sciences*, 67 (9): 1409-1427.
- Carruthers, T.R., Kimoto, A., Powers, J., Kell, L., Butterworth, D., Lauretta, M. and Kitakado, T. 2015a. Structure and estimation framework for Atlantic bluefin tuna operating models. *Col. Vol. Sci. Pap. ICCAT*, 72(7): 1782-1795.
- Carruthers, T.R., Powers, J., Lauretta, M., Di Natale, A., Kell, L. 2015b. A summary of data to inform operating models in management strategy evaluation of Atlantic bluefin tuna. *Col. Vol. Sci. Pap. ICCAT*, 72(7): 1796-1807.
- 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: experiences in the South African pelagic fishery. *Rev. Fish. Biol. Fisher.* 8, 177-214.
- Davis, K., Kragt, M., Gelcich, S., Schilizzi, S., Pannell, D. 2015. Accounting for enforcement costs in the spatial allocation of marine zones. *Conserv. Biol.* 29(1): 226-237.
- Francis, R, I. C, C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aqua. Sci.* 68(6): 1124-1138.
- GBYP. 2014. ICCAT Atlantic wide research programme for Bluefin Tuna. Available online at: <http://www.iccat.int/GBYP/en/index.htm> [accessed October 2014]
- Goni N., Onandia I., Uranga J., Arregui I., Martinez U., Boyra G., Arrizabalaga H. and Santiago J. 2016. First acoustic survey for a fishery-independent abundance index of juvenile bluefin tunas in the Bay of Biscay. *Col. Vol. Sci. Pap. ICCAT*, 72(7): 1862-1876.
- Hilborn, R. 2002. The dark side of reference points. *Bull. Mar. Sci.* 70(2): 403-408.
- Ingram, G.W., Jr., D. Alvarez-Berastegui, P. Reglero, R. Balbín, A. García5, and F. Alemany. Indices of larval bluefin tuna (*Thunnus thynnus*) in the Western Mediterranean Sea (2001-2013). Document SCRS/2015/035 (withdrawn).
- McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. *Can. J. Fish. Aqua. Sci.* 54: 284-300.
- MRAG 2009. Review of Fish Sustainability Information Schemes. Marine Resources Assessment Group. [http://cels.uri.edu/urissi/docs/FSIG\\_Report.pdf](http://cels.uri.edu/urissi/docs/FSIG_Report.pdf) [accessed 3/9/2016]

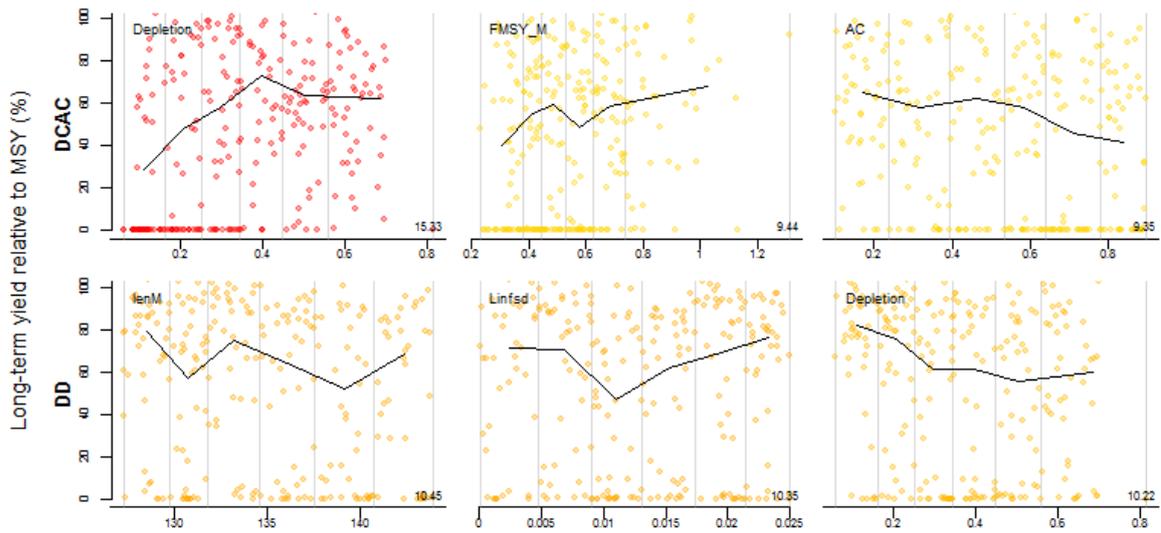
- Patrick, W.S., Spencer, P., Ormseth, O., Cope, J., Field, J., Kobayashi, D., Gedamke, T., Cortés, E., Bigelow, K., Overholtz, W., Link, J., Lawson, P., 2009. Use of productivity and susceptibility indices to determine stock vulnerability, with example applications to six U.S. fisheries. NOAA Fisheries, NOAA Technical Memorandum NMFS-F/SPO-101.
- Punt, A.E., Butterworth, D.S., de Moore, C. L., De Oliveira, J. A. A., Haddon, M. 2014. Management strategy evaluation: best practices. Fish and Fisheries. doi: 10.1111/faf.12104.



**Figure 1.** An example of an analysis of the value of better data. Each panel plots 300 points representing independent simulations. The value of bias or imprecision that was simulated is plotted on the x-axis, the relative yield obtained over each simulation is plotted on the y-axis. For both DCAC and delay difference (DD) management procedures, the three most important observation processes for determining yield are plotted. For example, ‘Cbias’ is the bias in historical catch observations where the value of 1 on the x-axis represent unbiased reporting of annual catches. It is clear from this plot that the delay-difference assessment (DD) is more vulnerable to catch over-reporting than DCAC and obtains lower yields on average, when catch is overestimated (e.g. Cbias>1.5). FMSY\_Mbias is the bias in the simulated level of FMSY relative to natural mortality rate  $M$ . Dbias is bias in observations of stock depletion (spawning biomass relative to unfished levels). lenMbias is bias in the observed length at 50% maturity. Isd is the lognormal error in observations of a relative abundance index.



**Figure 2.** An example of an analysis of the value of additional data. A simple trade-off plot shows the expected probability of not overfishing and the expected long term yield (as a fraction of FMSY yield) of two management procedures. The delay-difference model (DD) is unavailable due to a lack of information regarding growth. With these data this MP may be used instead of DCAC and the expected gain in yield is around 20%.



**Figure 3.** An example of an analysis of the cost of uncertainty. Each point represents one of 300 simulations. The level of various operating model parameters is plotted on the x-axis, the yield obtained is plotted on the y-axis. The top left panel shows that the yields obtained by the DCAC MP is strongly determined by the starting level of stock depletion (spawning biomass relative to unfished levels). When this is below 0.2 (20% of unfished biomass) DCAC obtains less than half the yield of the delay-difference (DD) MP. FMSY\_M is the simulated ratio of FMSY to natural mortality rate. AC is lag-1 recruitment autocorrelation. lenM is the length at 50% maturity. Linfsd is the interannual variation in maximum length, Linf.