



Fisheries and Oceans
Canada

Pêches et Océans
Canada

Science

Sciences

C S A S

Canadian Science Advisory Secretariat

Proceedings Series 2009/058

S C C S

Secrétariat canadien de consultation scientifique

Compte rendu 2009/058

**CEMAM Workshop on Bayesian
Modelling of Marine
Mammal Population Dynamics**

**27-29 March 2007
Freshwater Institute
501 University Crescent,
Winnipeg, MB, Canada**

Chairperson: Pierre R. Richard

**Atelier du CEMAM sur la modélisation
bayésienne de la dynamique des
populations de mammifères marins**

**27-29 mars 2007
Institut des eaux douces
501, University Crescent,
Winnipeg, MB, Canada**

Président : Pierre R. Richard

Fisheries and Oceans / Pêches et Océans
Arctic Aquatic Research Division /
Division de la recherche aquatique dans l'Arctique
501 University Crescent
Winnipeg, Manitoba, R3T 2N6
Pierre.Richard@dfo-mpo.gc.ca

March 2010

Mars 2010

Foreword

The purpose of these Proceedings is to document the activities and key discussions of the meeting. The Proceedings include research recommendations, uncertainties, and the rationale for decisions made by the meeting. Proceedings also document when data, analyses or interpretations were reviewed and rejected on scientific grounds, including the reason(s) for rejection. As such, interpretations and opinions presented in this report individually may be factually incorrect or misleading, but are included to record as faithfully as possible what was considered at the meeting. No statements are to be taken as reflecting the conclusions of the meeting unless they are clearly identified as such. Moreover, further review may result in a change of conclusions where additional information was identified as relevant to the topics being considered, but not available in the timeframe of the meeting. In the rare case when there are formal dissenting views, these are also archived as Annexes to the Proceedings.

This workshop was not carried out as a formal Fisheries and Oceans Canada (DFO) Science Advisory process; however, it is being documented in the Canadian Science Advisory Secretariat's (CSAS) Proceedings series as it presents some topics of interest related to the advisory process.

Avant-propos

Le présent compte rendu a pour but de documenter les principales activités et discussions qui ont eu lieu au cours de la réunion. Il contient des recommandations sur les recherches à effectuer, traite des incertitudes et expose les motifs ayant mené à la prise de décisions pendant la réunion. En outre, il fait état de données, d'analyses ou d'interprétations passées en revue et rejetées pour des raisons scientifiques, en donnant la raison du rejet. Bien que les interprétations et les opinions contenus dans le présent rapport puissent être inexacts ou propres à induire en erreur, ils sont quand même reproduits aussi fidèlement que possible afin de refléter les échanges tenus au cours de la réunion. Ainsi, aucune partie de ce rapport ne doit être considéré en tant que reflet des conclusions de la réunion, à moins d'indication précise en ce sens. De plus, un examen ultérieur de la question pourrait entraîner des changements aux conclusions, notamment si l'information supplémentaire pertinente, non disponible au moment de la réunion, est fournie par la suite. Finalement, dans les rares cas où des opinions divergentes sont exprimées officiellement, celles-ci sont également consignées dans les annexes du compte rendu.

Le présent atelier n'a pas été tenu dans le cadre officiel du processus des avis scientifiques du ministère des Pêches et des Océans (MPO). Celui-ci est toutefois documenté dans la série des comptes rendus du Secrétariat canadien de consultation scientifique (SCCS), car il couvre certains sujets en lien avec le processus des avis.

**CEMAM Workshop on Bayesian
Modelling of Marine
Mammal Population Dynamics**

**27-29 March 2007
Freshwater Institute
501 University Crescent,
Winnipeg, MB, Canada**

Chairperson: Pierre R. Richard

**Atelier du CEMAM sur la modélisation
bayésienne de la dynamique des
populations de mammifères marins**

**27-29 mars 2007
Institut des eaux douces
501, University Crescent,
Winnipeg, MB, Canada**

Président : Pierre R. Richard

Fisheries and Oceans / Pêches et Océans
Arctic Aquatic Research Division /
Division de la recherche aquatique dans l'Arctique
501 University Crescent
Winnipeg, Manitoba, R3T 2N6
Pierre.Richard@dfo-mpo.gc.ca

March 2010

Mars 2010

© Her Majesty the Queen in Right of Canada, 2010
© Sa Majesté la Reine du Chef du Canada, 2010

ISSN 1701-1272 (Printed / Imprimé)
ISSN 1701-1280 (Online / En ligne)

Published and available free from:
Une publication gratuite de :

Fisheries and Oceans Canada / Pêches et Océans Canada
Canadian Science Advisory Secretariat / Secrétariat canadien de consultation scientifique
200, rue Kent Street
Ottawa, Ontario
K1A 0E6

<http://www.dfo-mpo.gc.ca/csas/>

CSAS@DFO-MPO.GC.CA



Correct citation for this publication:
On doit citer cette publication comme suit :

DFO. 2010. CEMAM Workshop on Bayesian Modelling of Marine Mammal Population Dynamics; 27-29 March 2007.
DFO Can. Sci. Advis. Sec. Proceed. Ser. 2009/058.

MPO. 2010. Atelier du CEMAM sur la modélisation bayésienne de la dynamique des populations de mammifères marins ; 27-29 mars 2007. Secr. can. de consult. sci. du MPO, Compte rendu 2009/058.

SUMMARY

The Center of Expertise in Marine Mammalogy organised a workshop to bring together DFO and outside experts and review approaches being used to model marine mammal population dynamics, the current state of development of Bayesian modelling, and to identify challenges and issues related to this approach in its use on Canadian marine mammals. The CEMAM Workshop on Bayesian Modelling of Marine Mammal Population Dynamics was held on 27-29 March 2007 at the Freshwater Institute. This report summarizes the conclusions of the workshop and, in appendices, gives extended abstracts of the workshop presentations and a list of references on Bayesian modelling.

SOMMAIRE

Le Centre d'expertise sur les mammifères marins (CEMAM) a organisé un atelier regroupant des experts du MPO et de l'extérieur dans le but d'examiner les approches utilisées pour modéliser la dynamique des populations de mammifères marins et l'état d'avancement de la modélisation bayésienne ainsi que pour relever les difficultés et les problèmes reliés à l'application de cette approche aux mammifères marins au Canada. L'atelier du CEMAM sur la modélisation bayésienne de la dynamique des populations de mammifères marins a eu lieu du 27 au 29 mars 2007 à l'Institut des eaux douces. Le présent rapport résume les conclusions formulées pendant l'atelier et contient, en annexe, les résumés complets des exposés donnés au cours de l'atelier ainsi qu'une liste de références sur la modélisation bayésienne.

INTRODUCTION

Over the past decades, the field of ecology and resource management has seen increasing use of Bayesian modelling (Dixon and Ellison 1996; Ellison 2004; Clarke 2005). This trend is also noticeable in marine mammal science (Givens *et al.* 1995; Wade 1999, 2002; Ver Hoef and Frost 2003; Hobbs *et al.* 2006). In 2006, the DFO's Center of Expertise on Marine Mammalogy (CEMAM) acknowledged this methodological trend and recognised the value of these methods in the work of DFO marine mammalogists, in particular with respect to population monitoring.

Bayesian modelling is an approach that incorporates the uncertainty associated with data and prior knowledge on model parameters used to estimate abundance and trend of marine mammals. Currently, a number of DFO projects are underway to develop population models for marine mammals using a Bayesian framework (e.g. harp seals, grey seals, beluga whales). However, these projects would benefit from discussions with other experts in the field before they are used to generate scientific advice. Furthermore, DFO marine mammalogists not familiar with these methods would benefit from learning from other researchers experience using a Bayesian modelling framework for population modelling.

For these reasons, CEMAM organised a workshop to bring together DFO and outside experts to review approaches being used, the current state of development of Bayesian modelling, and to identify challenges and issues related to this approach in its use on Canadian marine mammals.

INTRODUCTION

Depuis quelques décennies, la modélisation bayésienne gagne en popularité dans les domaines de l'écologie et de la gestion des ressources (Dixon et Ellison, 1996; Ellison, 2004; Clarke, 2005). On remarque également cette tendance du côté de la science des mammifères marins (Givens *et al.*, 1995; Wade, 1999, 2002; Ver Hoef et Frost, 2003; Hobbs *et al.*, 2006). En 2006, le Centre d'expertise sur les mammifères marins (CEMAM) du MPO reconnaissait cette tendance méthodologique ainsi que la valeur de ces méthodes dans les travaux des experts des mammifères marins du MPO, surtout en ce qui concerne la surveillance des populations.

La modélisation bayésienne est une approche qui incorpore l'incertitude associée aux données et les connaissances *a priori* concernant les paramètres des modèles utilisés pour estimer l'abondance des mammifères marins et les tendances connexes. Actuellement, le MPO mène un certain nombre de projets pour élaborer des modèles des populations de mammifères marins en ayant recours à un cadre bayésien (p. ex. phoques du Groenland, phoques gris, bélugas). Cependant, les responsables de ces projets pourraient tirer profit de discussions avec d'autres experts du domaine avant d'utiliser les résultats de ces modèles pour produire des avis scientifiques. En outre, les experts en mammifères marins du MPO, qui ne sont pas très familiers avec ces méthodes, pourraient tirer profit de l'expérience d'autres scientifiques qui utilisent un cadre de modélisation bayésien pour modéliser des populations.

Le CEMAM a donc organisé un atelier regroupant des experts du MPO et de l'extérieur dans le but d'examiner les approches utilisées pour modéliser la dynamique des populations de mammifères marins et l'état d'avancement de la modélisation bayésienne ainsi que pour relever les difficultés et les enjeux liés à l'application de cette approche aux

mammifères marins au Canada.

WORKSHOP OBJECTIVES AND SUMMARY

The CEMAM Workshop on Bayesian Modelling of Marine Mammal Population Dynamics was held on 27-29 March 2007 at the Freshwater Institute. Invited speakers were: Alf Harbitz (IMR, Norway), Rod Hobbs (NMML, USA), Michael Kingsley (GNI, Grønland), Geoff, Evans (DFO Newfoundland) and Georges Watters (SWFSC, USA). In addition, ten DFO research scientists and biologists participated in the workshop and the meeting was open to observers from the University of Manitoba (Appendix I).

Objectives of the workshop were:

- To review the state of science on Bayesian modelling of population dynamics with emphasis on marine mammal population dynamic models.
- To identify statistical issues with model parameter estimation and produce guidelines for such modelling (choice and influence of priors, data load, convergence and autocorrelation issues, etc.).
- To review presentations of a few working models that have been implemented in software and problems in reaching convergence.
- To review software or programming issues and get insights from actual users/programmers on computing issues.

The invited speakers were told to keep in mind that the audience is mixed in its knowledge of Bayesian methods, so to keep language simple or explain complex terms. The audience was encouraged to ask questions of clarification but to keep them short and defer

OBJECTIFS ET RÉSUMÉ DE L'ATELIER

L'atelier du CEMAM sur la modélisation bayésienne de la dynamique des populations de mammifères marins a eu lieu du 27 au 29 mars 2007 à l'Institut des eaux douces. Les conférenciers invités étaient : Alf Harbitz (IMR, Norvège), Rod Hobbs (NMML, États-Unis), Michael Kingsley (GNI, Groenland), Geoff Evans (MPO, Terre-Neuve) et Georges Watters (SWFSC, États-Unis). Finalement, dix scientifiques et biologistes du MPO ont participé à l'atelier, et des observateurs de l'Université du Manitoba y ont été invités (annexe I).

Les objectifs de l'atelier sont les suivants.

- Passer en revue l'état de la science relative à la modélisation bayésienne de la dynamique des populations, notamment en ce qui concerne les modèles de la dynamique des populations de mammifères marins.
- Relever les problèmes statistiques associés à l'estimation des paramètres des modèles et élaborer des lignes directrices pour cette modélisation (choix et incidence des données *a priori*, volume de données, convergence et problèmes d'autocorrélation, etc.).
- Passer en revue des exposés sur quelques modèles fonctionnels qui ont été mis en application dans le logiciel ainsi que les problèmes concernant l'atteinte de convergence.
- Passer en revue les problèmes de logiciel ou de programmation et obtenir des précisions de la part d'utilisateurs/programmeurs sur les problèmes de calcul.

On a invité les conférenciers à ne pas oublier que l'auditoire est composé d'un éventail de personnes aux connaissances variées des méthodes bayésiennes et qu'il convenait d'utiliser un langage simple et d'expliquer les termes complexes. On a également invité

debate for the discussion period.

This report summarizes workshop conclusions on Bayesian modelling generated by the discussion amongst participants following presentations and in the final discussion period. Appendices I and II give the list of participants and the workshop agenda. Extended abstracts of the presentations and discussion specific to the models are provided in Appendix III. In addition, a list of references relevant to Bayesian statistics and modelling (in addition to the ones cited below) has been compiled in Appendix IV. This annotated list includes introductory texts for those who are seeking primers on Bayesian modelling and some of the software in common use.

WORKSHOP CONCLUSIONS ON BAYESIAN POPULATION MODELLING

Following presentations and discussion of each model, there was a general discussion on Bayesian population modelling to distil what we had learned and address workshop objectives. Participants chose to order the discussion in a logical sequence, from model choice to convergence and other issues:

- Model choice
- Software and programming issues
- Choice of priors
- Autocorrelation
- Convergence
- Sufficiency of data
- Other issues

l'auditoire à poser de courtes questions pour obtenir des précisions et d'attendre à la période de discussion pour en débattre davantage.

Le présent rapport résume les conclusions formulées au cours de l'atelier sur la modélisation bayésienne, à la suite des discussions qui ont suivi les exposés et la période de discussion finale. Les annexes I et II présentent la liste des participants et l'ordre du jour de l'atelier. Des résumés complets des exposés et des discussions sur les modèles se trouvent à l'annexe III. En outre, l'annexe IV donne une liste de références sur la statistique et la modélisation bayésienne (en plus de celles énumérées ci-après). Cette liste annotée comprend les textes d'introduction, pour ceux qui souhaitent en savoir plus sur les fondements de la modélisation bayésienne ainsi que l'information sur le logiciel couramment utilisé.

CONCLUSIONS DE L'ATELIER SUR LA MODÉLISATION BAYÉSIENNE DES POPULATIONS

Après les exposés et les discussions sur chaque modèle, on a tenu une discussion générale sur la modélisation bayésienne des populations pour résumer ce que l'on avait appris et pour répondre aux objectifs de l'atelier. Les participants ont décidé de tenir la discussion dans un ordre logique : du choix du modèle, en passant par la convergence, jusqu'à d'autres enjeux.

- Choix du modèle
- Problèmes concernant le logiciel et la programmation
- Choix des données *a priori*
- Autocorrélation
- Convergence
- Suffisance des données
- Autres enjeux

MODEL CHOICE

An important consideration in Bayesian modelling is to have a model that both represents the population or process and can answer the question that is posed. Bayesian methods can give results despite the fact that the model fails to represent the process adequately [this is of course also true of non-Bayesian models]. It is recommended that a modeller learn about the problem at hand by progressing from simple deterministic modelling to moderate process error modelling (likelihood methods) before moving on to a full Bayesian model. This step-by-step approach gives one a better understanding of the added complexity of the model as it develops and improves interpretation of the results of the final Bayesian analysis. Frequently, several models are plausible. In such cases, model choice can be aided by the DIC criterion or by Bayes factors.

BAYESIAN SOFTWARE AND PROGRAMMING ISSUES

The Bayesian modelling method most used by workshop presenters is the Markov Chain Monte Carlo (MCMC) Gibbs sampling method. For that approach to Bayesian modelling, WinBUGs or OpenBUGS are excellent programs for people that are new to Bayesian modelling. They have a serious MCMC Gibbs sampler with many strong features for diagnostics and graphics. They are relatively easy to use and are free. The JAGS program is another useful free Gibbs sampler. It can compile models that don't compile in WinBUGS or OpenBUGS. It is faster and has additional functions, such as data simulation for data sufficiency checks. JAGS is not as easy to use. It has no Windows GUI and it does not allow cyclic walks as BUGS programs do. ADModel Builder does likelihood and Bayesian (MCMC) modelling and is extremely fast but it is very difficult to specify a model [ADMB used to be expensive

CHOIX DU MODÈLE

Dans la modélisation bayésienne, il importe que le modèle à la fois représente la population ou le processus et réponde à la question soulevée. Les méthodes bayésiennes peuvent donner des résultats, même si le modèle ne représente pas le processus adéquatement [cela est également vrai pour les modèles non bayésiens, bien entendu]. On recommande que le modélisateur prenne connaissance du problème à résoudre en passant d'une modélisation déterministe simple à une modélisation probabiliste modérée (avec erreur dans les variables), avant d'en arriver à un modèle bayésien complet. Cette démarche progressive facilite la compréhension de la complexité grandissante du modèle au fur et à mesure de son développement et améliore l'interprétation des résultats de l'analyse bayésienne finale. Souvent, plusieurs modèles sont plausibles. Dans de tels cas, le choix du modèle peut être facilité à l'aide du critère d'information de déviance (DIC, ou Deviance Information Criterion) ou de facteurs de Bayes.

PROBLÈMES CONCERNANT LE LOGICIEL ET LA PROGRAMMATION

La méthode de modélisation bayésienne la plus utilisée par les présentateurs est la méthode d'échantillonnage de Gibbs, à savoir la méthode Monte Carlo par chaîne de Markov (MCMC). Avec cette démarche de modélisation bayésienne, WinBUGs ou OpenBUGS sont d'excellents programmes pour les néophytes de la modélisation bayésienne. Ils offrent un excellent échantillonneur de Gibbs MCMC doté de nombreuses caractéristiques fort utiles pour le diagnostic et les graphiques. Ils sont relativement faciles à utiliser et sont gratuits. Le programme JAGS est un autre échantillonneur de Gibbs utile et gratuit. Il peut compiler des modèles qui ne se compilent pas avec WinBUGS ou OpenBUGS. Il est plus rapide et offre des fonctions supplémentaires, telles que la simulation de données pour les vérifications de la suffisance des données. Le programme JAGS est cependant difficile à

to buy but it is now freeware supported by a non-profit user group]. The mathematical programming package MatLab also has functions for MCMC. It requires the user to learn the language and is very expensive.

The Sampling Importance Resampling (SIR) approach covered by R. Hobbs is more efficient than MCMC. It keeps only important samples and it does not require starting values. There are no canned programs for SIR, so it does require knowledge in some programming language, which will be a deterrent to people unfamiliar with programming.

CHOICE OF PRIORS

It is preferable to use non-informative priors at first to learn from results. One can then see if data within the model data set contain information that will allow the posteriors to update the priors of the model parameters. Caution should be used when considering priors that are too informative, as there may be no updating. It is good practice to use priors that are informative only to the extent necessary to run the program efficiently. Scaling of informative priors is also important. For models to be used in management decision-making it is preferred that informed priors be supported by objective data as subjective priors will depend on agreement amongst experts. Finally, it is important to find means of quantifying qualitative priors.

utiliser. Il n'a pas d'interface graphique Windows et ne permet pas les trajets cycliques comme le permettent les programmes BUGS. ADModel Builder, qui établit la probabilité et effectue la modélisation bayésienne (MCMC), est extrêmement rapide, mais il est très difficile de préciser un modèle [ADMB était dispendieux à l'achat, mais il est maintenant gratuit grâce au soutien d'un groupe d'utilisateurs sans but lucratif]. Le progiciel mathématique MatLab présente également des fonctions MCMC. Cependant, l'utilisateur doit apprendre le langage de ce progiciel, qui est très dispendieux.

L'approche de l'échantillonnage avec rééchantillonnage par importance (Sampling Importance Resampling, ou SIR) examinée par R. Hobbs est plus efficace que la méthode MCMC. Seuls les échantillons importants sont conservés et il n'exige pas de valeurs de départ. Comme il n'existe aucun programme de série pour l'approche SIR, il faut avoir une certaine connaissance du langage de programmation, ce qui dissuadera ceux qui sont peu familiers avec la programmation.

CHOIX DES DONNÉES A PRIORI

Il vaut mieux utiliser des données *a priori* non informatives en premier lieu pour apprendre des résultats. On peut alors voir si les données de l'ensemble de données du modèle contiennent l'information qui permettra aux données *a posteriori* de mettre à jour les données *a priori* des paramètres du modèle. Il faut faire preuve de circonspection lorsqu'on utilise des données *a priori* trop informatives, car il pourrait ne pas y avoir de mise à jour. La bonne pratique consiste à utiliser des données *a priori* qui sont informatives uniquement dans la mesure nécessaire pour exécuter le programme efficacement. Il est également important de mettre les données *a priori* informatives à l'échelle. Pour les modèles destinés à la prise de décisions de gestion, on préfère que les données *a priori* soient étayées par des données objectives, puisque les données *a priori* subjectives dépendront d'un accord entre des spécialistes. Finalement, il est important de trouver des

moyens de quantifier les données *a priori* qualitatives.

AUTOCORRELATION

This is not a problem with the SIR algorithm because each trial is independent. And the algorithm retains only the important samples. The Gibbs sampler on the other hand, is a searching routine and each step in the search is based on and consequently correlated to the previous. In Gibbs sampling programs, autocorrelation is to be expected in any of these multi-parameter fitting routines. It can be dealt with by thinning every n samples. The choice of a thinning n must be done by trial and error: setting n and checking for reduction in auto-correlation. It is best to thin before sampling; otherwise some WinBUGS estimates for example may not be correct

CONVERGENCE

Lack of convergence can be a problem with data that has little information and weak (un-informative) priors. Convergence can also be assessed by looking at the sample trace of estimated parameters. If a trace shows trends (i.e: searching) rather than a random sampling behaviour then there is a problem of convergence. Convergence can be assessed by doing several runs and comparing them. If the different runs converge to the same estimates, then there is good convergence. If they do not, then there is a convergence problem. There are tools in WinBUGS/OpenBUGS to compare runs. There are also diagnostic tools in R and S-plus to check for convergence (CODA, BOA) but they can be misleading. These tools are somewhat adhoc empirical takes on diagnostics, which is strange for Bayesian analysis. Another approach is to use Q-Q plots to compare posteriors and priors. Re-parametization of the model can help convergence speed. While convergence is not

AUTOCORRÉLATION

Ce problème ne se pose pas avec l'algorithme SIR puisque chaque essai est indépendant. L'algorithme ne conserve que les échantillons importants. L'échantillonneur de Gibbs est quant à lui une routine de recherche, et chaque étape de la recherche est fondée sur l'étape précédente et, par conséquent, corrélée avec celle-ci. Dans les programmes d'échantillonnage de Gibbs, on s'attend à ce qu'il y ait autocorrélation dans toutes ces routines d'ajustement de paramètres multiples. Cela peut être obtenu par amincissement à tous les échantillons n . Le choix d'intervalle d'amincissement n doit être fait par essais et erreurs : établissement de n et vérification de la réduction de l'autocorrélation. Il est préférable d'amincir avant d'échantillonner, sinon certaines estimations de WinBUGS, par exemple, peuvent être incorrectes.

CONVERGENCE

Le manque de convergence peut survenir lorsque les données reposent sur peu d'information et que les données *a priori* (non informatives) sont peu solides. La convergence peut également être évaluée avec la trace des échantillons des paramètres estimés. Si une trace montre des tendances (i.e.: recherche) plutôt qu'un comportement d'échantillonnage aléatoire, il y a alors un problème de convergence. On peut évaluer la convergence en faisant plusieurs passages du modèle et en les comparant. Si les divers passages convergent vers les mêmes estimations, c'est que la convergence est bonne. Dans le cas contraire, il y a un problème de convergence. WinBUGS/OpenBUGS offrent des outils pour comparer des passages. Il y a également des outils de diagnostic en R et S-plus pour vérifier la convergence (CODA, BOA), mais ils peuvent être trompeurs. Ces outils donnent en quelque sorte des clichés empiriques ponctuels pour le diagnostic, qui est

a problem with the SIR algorithm, it is necessary to test the number of trials that are run for the SIR sampling to determine the precision of the posteriors. Hobbs indicated that he typically used between 10^5 and 10^6 to insure 2 significant figure precision in the percentiles of the posteriors.

DATA SUFFICIENCY

The concept that Bayesian modelling is for data poor situations is debatable. There is little one can do even with a good dataset, if there is too much observation error and too much underlying process error. However, the advantage of the Bayesian approach is that it will provide a measure of the sufficiency of the data by comparison of the posterior to the prior distributions of the parameters of interest. More specific to population models, catch data alone cannot be used to estimate population size. The model must hinge on the likelihood of one or more population indices to allow convergence to an estimate.

OTHER ISSUES MENTIONED

When doing risk analyses based on Bayesian model outputs, the risk model should assess the decision criterion from the posterior distributions of parameter estimates. The use of a similarly-shaped function to the posterior distribution, derived from its mean and SD for example, will reduce the efficiency of the analyses and introduce more error. The analyst should use the parameter posterior samples to model parameter uncertainty. With WinBUGs or other similar software, it is easy to use those samples to project forward.

surprenant en analyse bayésienne. Une autre approche consiste à utiliser des graphiques Q-Q afin de comparer les données *a posteriori* et les *données a priori*. La reparamétrisation du modèle peut aider à améliorer la vitesse de convergence. Même si la convergence ne pose pas de problème avec l'algorithme SIR, il faut examiner le nombre d'essais qui sont effectués pour l'échantillonnage SIR afin de déterminer la précision des données *a posteriori*. Hobbs a indiqué qu'il utilisait d'ordinaire entre 10^5 et 10^6 pour obtenir une précision de deux chiffres significatifs dans les percentiles des données *a posteriori*.

SUFFISANCE DES DONNÉES

L'idée voulant que la modélisation bayésienne soit utilisée dans les situations où les données sont rares est discutable. On ne peut faire grand-chose, même avec un bon ensemble de données, lorsqu'il y a trop d'erreurs d'observation et trop d'erreurs de traitement sous-jacentes. Cependant, l'avantage de l'approche bayésienne est qu'elle donne une mesure de la suffisance des données en comparant les distributions postérieures et antérieures des paramètres d'intérêt. Dans le cas des modèles sur les populations, on ne peut utiliser des données sur les prises seules pour estimer la taille d'une population. Le modèle doit s'appuyer sur la probabilité d'un ou de plusieurs indices de la population pour permettre la convergence vers une estimation.

AUTRES ENJEUX MENTIONNÉS

Lorsqu'on effectue des analyses du risque en s'appuyant sur les résultats des modèles bayésiens, le modèle du risque doit évaluer le critère de décision à partir des distributions *a posteriori* des estimations de paramètre. L'utilisation d'une fonction de forme similaire à la distribution *a posteriori*, dérivée de sa moyenne et de son écart-type, par exemple, réduira l'efficacité des analyses et amènera plus d'erreurs. L'analyste doit utiliser les échantillons *a posteriori* du paramètre pour modéliser l'incertitude entourant le paramètre. Avec WinBUGs ou d'autres logiciels du genre, il est facile d'utiliser de tels échantillons pour

établir des projections.

GENERAL CONCLUSIONS

Should we be Bayesian?

Yes, Bayesian methods estimate probability correctly! For example, the posterior distribution of population size gives the true probability of population estimate being equal to a particular value, given the prior and the data. Also, one can't model process error using likelihood methods because process error is nested in sampling error. Bayesian modelling requires additional efforts but it has many rewards (estimates of many parameters).

If so, when and how should we use Bayesian modelling?

Bayesian modelling should be used when the model is moderate to complex, and when there is a hierarchical error structure. WinBUGS or OpenBUGS are best for beginners as well as many serious applications. But there are a number of software choice which allow for creativity (eg.: JAGS, MatLab, SIR programming) for those who seek to tackle complex problems. Bayesian methods should be used when there is a desire to link a decision based on probable risk directly linked to input data because it will allow simultaneous assessment of the value of the data in making the decision and the level of risk.

When should one not bother with Bayesian modelling?

It is not necessary to go through all the trouble of developing and solving a Bayesian model when the question and the model are simple and the adequacy of the data is not in

CONCLUSIONS GÉNÉRALES

Devons nous être bayésiens?

Oui, les méthodes bayésiennes donnent une estimation correcte de la probabilité! Par exemple, la distribution *a posteriori* de la taille de la population donne la probabilité véritable que l'estimation de la population soit égale à une valeur particulière, compte tenu des données *a priori* et des données. En outre, on ne peut pas modéliser les erreurs de traitement en utilisant des méthodes de vraisemblance du fait que les erreurs de traitement sont comprises dans l'erreur d'échantillonnage. La modélisation bayésienne exige des efforts supplémentaires, mais elle offre de nombreux avantages (estimation de nombreux paramètres).

Si oui, quand et comment devons-nous utiliser la modélisation bayésienne?

La modélisation bayésienne doit être utilisée lorsque le modèle est de modéré à complexe et quand il y a une structure d'erreurs hiérarchique. WinBUGS ou OpenBUGS conviennent parfaitement pour les débutants ainsi que pour de nombreuses applications plus complexes. Cependant, un certain nombre de logiciels permettent à ceux qui cherchent à s'attaquer à des problèmes complexes (par exemple, JAGS, MatLab, programmation SIR) d'exprimer toute leur créativité. Les méthodes bayésiennes doivent être utilisées lorsqu'on souhaite lier une décision fondée sur un risque probable directement lié aux données d'entrée, car elles permettent l'évaluation simultanée de la valeur des données dans la prise de décision et du niveau de risque.

Dans quelles circonstances doit-on laisser la modélisation bayésienne de côté?

Il n'est pas nécessaire d'effectuer tout le branle-bas relatif à l'élaboration et à l'exécution d'un modèle bayésien lorsque la question et le modèle sont simples et que

question. The reward is not worth the substantial effort of Bayesian modeling. Well-established empirical methods can do for such situations. It is worth checking, however, that the question and model really are simple. Often they are asserted to be simple for convenience. In linear regression, for example, Bayesian methods can deal with observations of variable precision with no extra effort.

ACKNOWLEDGMENTS

Garry Stenson wrote the initial proposal for this workshop. Garry and Mike Hammill helped convene the meeting. Unfortunately, on the week of the workshop, they both had to attend urgent seal issues and could not participate in the meeting. Many thanks to all the presenters who shared their unique experience with their working models with workshop participants and prepared extended abstracts for the report. We are grateful to Jack Lawson for acting as rapporteur on day 2 of the meeting.

l'adéquation des données ne soulève pas de problème. Les avantages tirés n'équivalent pas à la valeur de l'effort substantiel requis par la modélisation bayésienne. Des méthodes empiriques bien établies peuvent convenir en de telles situations. Il vaut cependant la peine de vérifier que la question et le modèle sont vraiment simples. Souvent, on affirme qu'ils sont simples pour des raisons de commodité. Dans la régression linéaire, par exemple, les méthodes bayésiennes peuvent traiter des observations de précision variable sans nécessiter d'effort supplémentaire.

REMERCIEMENTS

Garry Stenson a rédigé la proposition initiale concernant cet atelier. Garry et Mike Hammill ont contribué à l'organisation de la réunion. Malheureusement, la semaine au cours de laquelle l'atelier a eu lieu, ils ont tous deux dû s'occuper de questions urgentes concernant le phoque et n'ont pu prendre part à la réunion. Un grand merci à tous les présentateurs qui ont partagé leur expérience unique en matière de modèles fonctionnels avec les participants et qui ont préparé des comptes rendus pour le rapport. Nous tenons également à remercier Jack Lawson qui a agi en tant que rapporteur pendant la deuxième journée de la réunion.

REFERENCES

RÉFÉRENCES

- Clark, J.S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters*, (2005) 8: 2–14.
- Dixon, P. and A.M. Ellison. 1996. Introduction: Ecological Applications of Bayesian Inference. *Ecological Applications*: Vol. 6, No. 4, pp. 1034-1035. (see also the group of papers in this seminal collection / voir également les autres documents de cette collection d'intérêt)
- Ellison, A.M. 2004. Bayesian Inference in Ecology. *Ecology Letters* 7: 509-520.
- Givens, G.H., A.E. Raftery and J.E. Zeh. 1995. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using the BALEEN II model in a Bayesian synthesis framework. *Rep. int. Whal. Commn.* 45: 345-364.
- Hobbs, R.C. K.E.W. Sheldon, D.J. Vos, K.T. Goetz, and D.J. Rugh. 2006. Status review and extinction assessment of Cook Inlet belugas (*Delphinapterus leucas*). AFSC Processed Rep. 2006-6, 74 p. Alaska Fish. Sci. Cent., NOAA Natl Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.
- Ver Hoef, J.M., and K.J. Frost. 2003. A Bayesian hierarchical model for monitoring harbour seal changes in Prince William Sound, Alaska. *Environmental and Ecological Statistics* 10: 201-219.
- Wade. P.R. 1999. A comparison of statistical methods for fitting population models to data. In: Garner, G.W., S. C. Amstrup, J. L. Laake, B.E.J. Manly, L.L. McDonald, and D.G. Robertson (eds). *Marine Mammal Survey and Assessment Methods*. AA. Balkema Publ. Rotterdam. 287 p.
- Wade, P.R. 2002. A Bayesian stock assessment of the Eastern Pacific gray whale using abundance and harvest data from 1967-1996. *J. Cetacean Res. Manage.* 4(1): 85-98.

APPENDIX I

List of participants:

Invited speakers:

Geoff Evans, Department of Fisheries and Oceans, St.John's, NL, Canada
(Geoff.Evans@dfo-mpo.gc.ca)
Alf Harbitz, Institute of Marine Research (IMR), Tromsø, Norway
(alf.harbitz@imr.no)
Roderick Hobbs, National Marine Mammal Laboratory, NMFS Seattle, WA, USA
(Rod.Hobbs@noaa.gov)
Michael Kingsley, Grønlands Naturinstitut, Nuuk, Greenland
(mcsk@natur.gl)
George Watters, Southwest Fisheries Science Center, Pacific Grove, CA, USA
(George.Watters@noaa.gov)

Other Participants

Holly Cleator, DFO Winnipeg
Larry Dueck, DFO Winnipeg
Steve Ferguson, DFO Winnipeg
Jean-Francois Gosselin, DFO Mont-Joli
Lei Harris, DFO St.Andrews
Jeff Higdon, University of Manitoba
Jack Lawson, DFO St-John's
Kathleen Martin, DFO Winnipeg
Lisa Loseto, University of Manitoba
Pierre Richard, DFO Winnipeg
Tim Siferd, DFO Winnipet
Kent Smedbol, DFO St.Andrews

APPENDIX II

Workshop agenda

- 27 March 2007:

o Morning:

- Chair remarks
- Introduction to Bayesian Modelling: “A tutorial on Bayesian modelling of harp seals” speaker: Geoff Evans and Working model 1: “Fractions pregnant and population history of Newfoundland harp seals” speaker: Geoff Evans (Appendix III-B-1).

o Afternoon:

- Discussion of Model 1
- “An introduction to Bayesian modelling” (Appendix III-A) and Working model 2: “Shrimp estimation models” (Appendix III-B-2) speaker: Michael Kingsley
- Discussion of Model 2

-

- 28 March 2007:

o Morning:

- Working model 3: “Population dynamics of White and Barents Sea harp seals” speaker: Alf Harbitz
- 10:45 Discussion of Model 3

o Afternoon:

- Working model 4: “Population dynamics of (mostly) short-beaked common dolphins from (at least) three stocks in the eastern Pacific Ocean” speaker: George Watters (90 min).
- Discussion of Model 4
- Working model 5: “Bayesian population viability analysis (PVA) for the Cook Inlet, Alaska beluga population using a Sampling-Inference-Resampling (SIR) algorithm. Rod Hobbs

- 29 March 2007:

o Morning:

- General discussion on statistical and programming issues
- Workshop conclusions, and reporting plans

o Afternoon:

- Administrative matters

Appendix III

A) An introduction to Bayesian modelling (Michael Kingsley, Grønlands Naturinstitut)

To simplify, there are two statistical views of the world termed "Frequentist" and "Bayesian". The 'Likelihood' school falls somewhere in the middle.

What is statistics all about? Statistical induction. We know about Deduction, which is from a general rule, to make inferences about a particular case

Eg: "Epimenides is a Cretan; all Cretans are liars; Epimenides is a liar."

Induction is the other way round: from many particular cases, we attempt to infer a general rule. This is the common experience of human existence (nobody gives us the general rules) So, we run into a common problem: how many cases do we need, and how reliable is the general rule that we induce? Statistics is inductive reasoning in a quantitative form; it is the basis for science - because science is an organised quest for general statements about how the world works.

We need to weigh the claims of rival general statements.

Classical' or 'frequentist' statistics grew out of the school of repeated experiments with cards and dice, among other things. It pre-supposes that the world is —but our knowledge of it is uncertain. In other words, general rules exist and are revealed to us (uncertainly) by experiments. They produce probabilistic statements about the statements we make. For example, about the mean of a population: if we did this experiment many times; and analysed the data this way; and formulated our statement thus we could state that the mean lies between U and L" and we would be right 95% of the time'.

Frequentist statistics leads to: confidence intervals: We construct random intervals, based on our random experiments/observations. These random intervals may, or may not, contain the true value but the true value is there, regardless of our intervals. We hope analysis allows us:

- to make statements about our confidence intervals, confidence levels associated with definite statements, to quantitative tests of hypotheses:

"If these means were equal, we would have observed so large a difference in only 1 of 100 experiments carried out like this one."

We say these means are different, with confidence in what we say.

- to experimental design

Frequentist thinking lends itself to experimental design because experimental design considers the hypothetical samples, experimental design has hypothesis testing and error rates in mind,

- to statistical control theory and design of quality control schemes

Example: in 115 births, 60 boys, 55 girls. (unbiased) sex ratio estimate is $60/115 = 0.522$ because: if we have a lot of samples of births, and we repeatedly calculate the sex ratio this way;

the mean of our calculated values converges on the true value; confidence if the true sex ratio was bigger than 60.16%, we would have so few as, or fewer than, 60 boys in only 5% of many such samples; if the true sex ratio was less than 44.96%, we would have so many as, or more than, 60 boys in only 5% of many such samples; we have 90% confidence that the true value is within the interval 45.0 to 60.2%; but to make that statement, we had to consider the probability of many outcomes that we did not observe. (0 boys, 1, 2, 3, &c.) because we consider many samples, our calculations include the many possible birth orders; (binomial distribution)

Hypothesis test

If the true average sex ratio was 64%, and we took a sample of 115 births many times, we would observe 60 boys or fewer in only 0.6% of such samples; we reject the suggestion ('hypothesis') that the sex ratio is 64% (or more) 'at the 1% level'; we are 99% sure that our decision is the right one.

Likelihood inference

Likelihoods are relative; only one sample—60 boys, 55 girls; how likely is it? $L = k \cdot p^{60} \cdot (1-p)^{55}$; Maximum Likelihood: the sex ratio (p) is 52.2%; not because it has properties related to many such samples—but because it maximises the likelihood of this sample.

Confidence?

Likelihood support intervals: find values that reduce the (log) likelihood by a given amount; (without explanation) a 1.353 l.s.i. corresponds to a 90% confidence interval; for this sample, such an interval is 44.52% to 59.76%; sex ratios outside this interval reduce—too much—the likelihood of this sample relative to its greatest possible value; likelihood calculations do not include the many possible birth orders for the sample; only one sample; therefore only on order; all orders have the same likelihood; so order is irrelevant;

Bayesian statistics is different

Classical: the world is, but we don't know about it. The true value is there; we construct a confidence interval for an estimate; in such a way that it has a specified chance of containing the true value; probability statements about the interval;

Bayesian: the world isn't—but it might be. The value has a probability distribution, just like anything else; probability statements about the true value;

Bayesian statistics accepts probabilistic statements about the world. It produces probability distributions for the values of parameters that define how the world is. It doesn't produce hypothesis tests in quite the same way. How does it do this?

Bayes' Theorem is a restatement of a simple truism in probability:

$$p(W \& O) = p(W) \cdot p(O|W)$$

$$p(O \& W) = p(O) \cdot p(W|O);$$

$$p(O) \cdot p(W|O) = p(W) \cdot p(O|W)$$

Consequently,

$$p(W|O) = p(W) \cdot p(O|W) / p(O)$$

which is Bayes theorem. It essentially states that the probability that the world is like so, given the observations we (have just) made $p(W|O)$ is equal to the probability that the world was like that anyway $p(W)$ multiplied by the probability of making those observations if the world was like that $p(O|W)$ and divided by the probability of making those observations anyway $p(O)$. $p(O)$ is the probability of making just those observations but that probability is equal to:

$$\int p(W) \cdot p(O|W) \cdot dW$$

It is the sum, over all states of the world, of the probability of the observations given that state of the world, multiplied by the (prior) probability of that state of the world!

So the bottom line of the Bayesian equation is the integral of the top line, the expression for $p(W|O)$ is a valid probability distribution (integrates to unity).

In practice, $p(O)$ is treated as a normalising constant; given that $\int p(W|O) = 1$,

$$p(W|O) = p(W) \cdot p(O|W) / \int (p(W) \cdot p(O|W)) dW$$

or more simply:

$$p(W|O) = k \cdot p(W) \cdot p(O|W)$$

where:

- $p(W)$ is the prior (distribution for W). This is a (quantitative) statement of what we knew about the world, before we made—independent of—the observations we just made;
- $p(O|W)$ is the likelihood of the observations, conditional on W ; and
- $p(W|O)$ is the posterior distribution for W , where the prior distribution has been updated by the information contained in the observations.

The major difference between Bayesian and Frequentist statistics is that Bayesian thinking is symmetrical: the World (W) and the Observations (O) get similar treatment in Bayes' theorem.

In practice, the world under study (W) consists of some set of variables (parameters); $p(W)$ (the prior) may (usually does) consist of separate univariate distributions of the individual variables, or multivariate distributions of sets of variables, or a mixture. $p(W|O)$ (the posterior) comprises one monstrous multivariate distribution of all the parameters we are considering, but we typically mostly look at the univariate marginal distributions of individual variables (caution—correlations!).

When we make observations in daily life, we use them to change our thinking about a lot of things; Bayesian analysis can use observations to update priors simultaneously on many variables, even on more variables than we have measurements. So the updating is not independent, results may be correlated. Caution is appropriate in using/interpreting Bayesian results.

In Bayesian analysis, priors are both a strength and a weakness. They are a strength because they represent our previous knowledge. They enable us to use this set of observations to update it. They are a weakness, because they must be included. They deceive us into

importing knowledge we haven't got. Discussion about priors is a major feature of the development of Bayesian methods, both in general and in particular.

Why does Bayesian statistics work? Bayesian statistics work because that's the way our minds work. We live with a probabilistic world-view (prior). We can accept probabilistic statements about the state of the world, more easily than confidence statements about the statements we make. Our minds work in a Bayesian way! It is a symmetrical formulation. Our prior knowledge and recent observations get equal treatment. There is the possibility of simultaneously considering many different data sources. There is no need to consider the hypothetical many repetitions of sampling or experimentation. Results are conditioned on one sample!

So why hasn't it been used so much before? The philosophical rigour of frequentist statistics is partly to blame but also the denominator of Bayes theorem ($\int (p(W) \cdot p(O|W))dW$) is terribly heavy on computing. But, also, it is difficult to formulate a (quantitative) prior distribution that accurately reflects what we (qualitatively) think we know. The prior is perhaps qualitative. There is a need to generate priors that are appropriately informative or uninformative depending on the case at hand. Bayes' idea of a continual updating probability versus one-shot analyses was not well received.

What did Reverend Bayes really mean? Perhaps he meant that we have an unstated, permanent, personal, qualitative prior, that is continually updated with new observations; where observations come a few at a time. What we do now with Bayesian statistics is assemble all the observations, construct synthetic priors, run all the observations at once; and then do it all again next time!! This last approach leaves almost no capability for putting this year's posterior in as next year's prior.

Discussion of presentation

This presentation gave an overview of Bayesian thinking and Bayesian modelling. Discussion was deferred to the working models in the subsequent presentations and in the conclusions of the workshop.

B) Bayesian modelling examples

This section summarizes the presentations of five invited speakers on their Bayesian population models. Four were marine mammal population models and the fifth (Kingsley) is a shrimp model which could be implemented on a marine mammal population with similar data. An attempt at doing so on Cumberland Sound belugas was presented but it was not deemed successful for reasons discussed below. The models are given here in the sequence that they were presented at the workshop.

B-2: Working model 1: A tutorial introduction to Bayesian modelling of harp seals. Geoffrey Evans, DFO St John's.

What beliefs are warranted by the evidence?

This is a sensible question, related to needs of management; the mathematics called probability theory helps one address it. I'm not interested in whether this is what probability "really" means (I'm not sure such a question makes sense.)

In fact the question as posed can't be answered. So the trick is to pick a nearby, proxy question:

What adjustments to previous beliefs are warranted by the evidence?

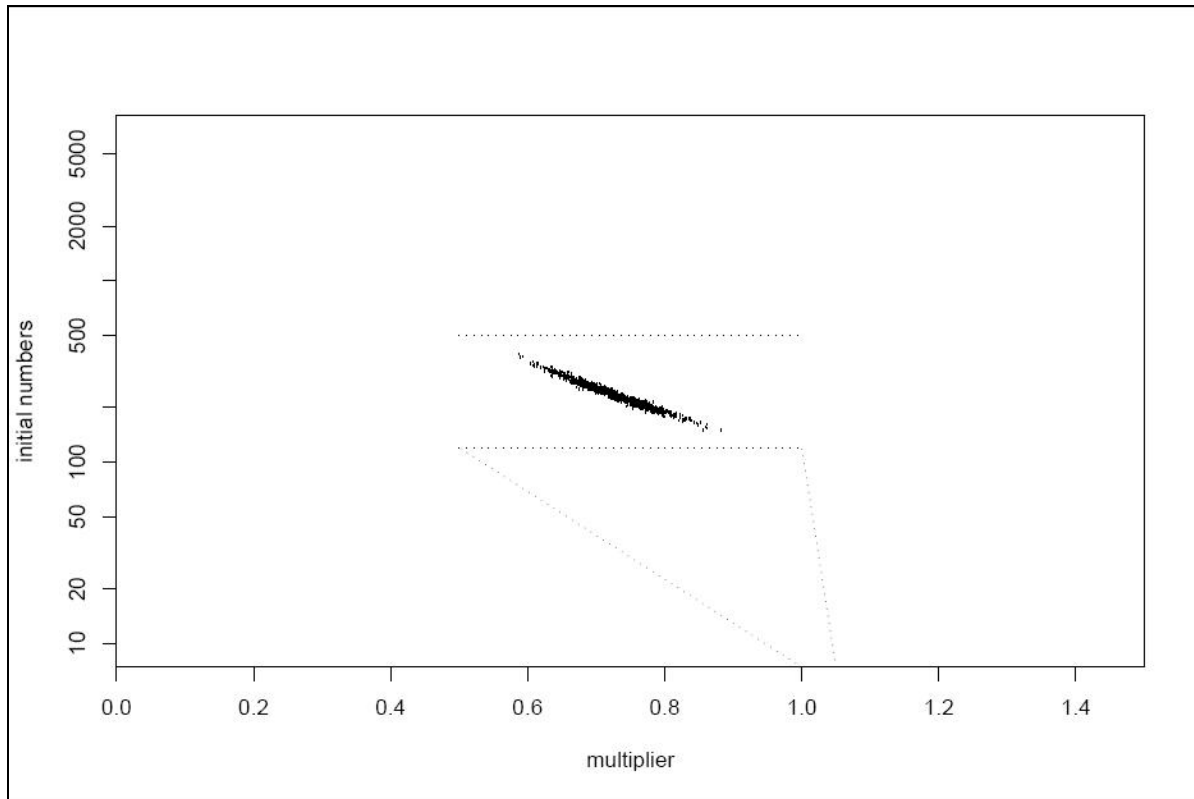
This question can be answered, and in many instances we can persuade ourselves that the answer we get is close enough to the answer to the original question.

Exponential growth of pups

I start with a toy model that nobody believes, to illustrate the way of thinking. We suppose that any year's population of pups is a constant multiple of the previous year's: the log of pup numbers is a straight line function of time. We start with a belief that the multiple could equally be anywhere between its minimum possible value of 0 and its maximum possible value of 1.5 (all pups survive and all females reproduce every year), and that the log of numbers in the year we choose as starting year could equally be anywhere between 3 and 20 (numbers between 20 and 22,000.) We seek an adjusted belief for what these two parameters could be, based on a few pup surveys.

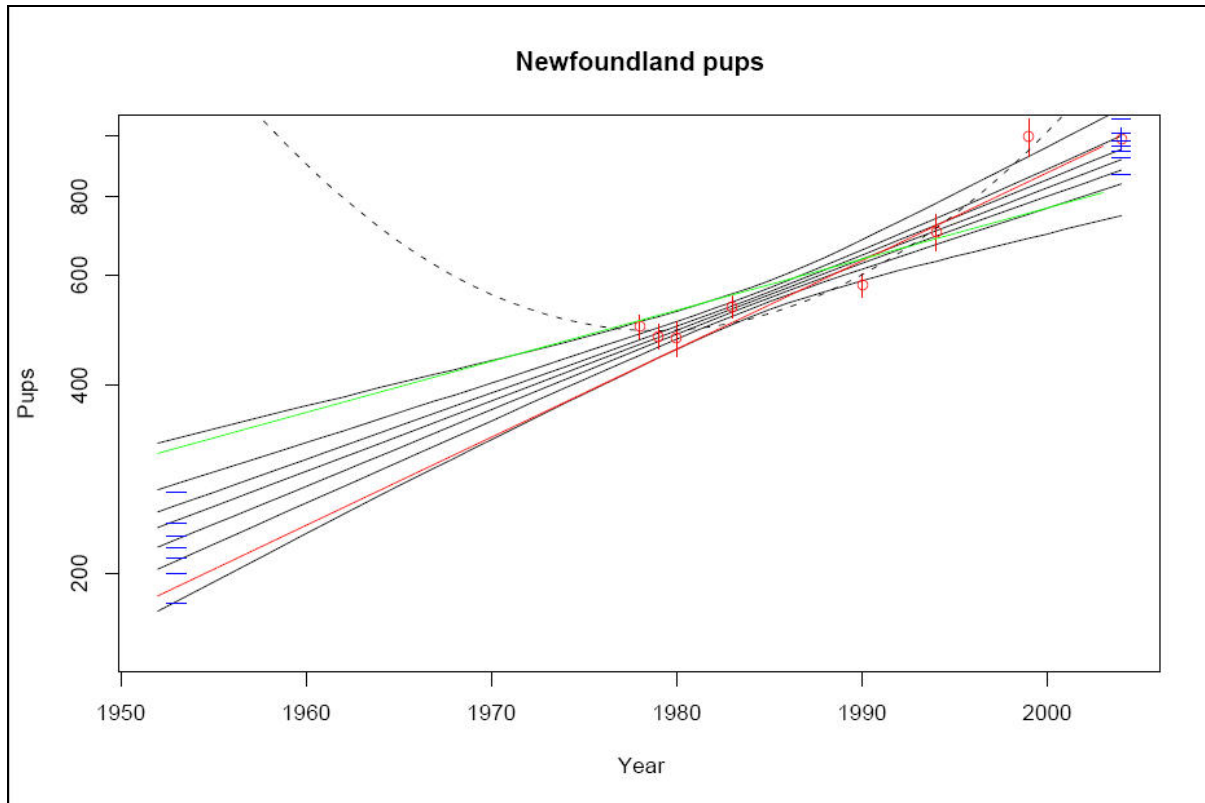
It's easy to work out the relative probability of two pairs of parameter values, just by dividing one instance of Bayes rule by another. All the work in Bayesian analysis comes when we seek absolute probability densities, which means working out the integral normalizing constant in the denominator. The method most often used is MCMC, which can conveniently be regarded as magic for now, which lets us draw random samples from the adjusted (posterior) probability distribution for parameter values. There is specialized software for doing this without having to understand too much; I use JAGS.

First, let's address the issue of whether answering the question "What adjustments to beliefs are warranted?" is close to answering the question "What beliefs are warranted?" Figure 1 shows the posterior distribution for the parameters. This distribution occupies only a tiny fraction of the prior space (which is indicated by the range of the plot axes). The posterior range of the multiplier has to be magnified to be visible at all. So it's easy to believe that any "uninformative" or "vague" prior would lead to about the same posterior.



Notice also that the estimates of the two parameters are strongly correlated, which makes sense but also makes trouble for the Gibbs sampler. One possible strategy is to do PCA on a preliminary MCMC run and choose as auxiliary random variables distances along the principal axes. This improves the chance of taking uncorrelated random samples from the posterior, and in practice greatly improves the chance of convergence.

As well as posterior distributions for the parameters, we also have them for any function of the parameters we care to compute. In particular we can examine our beliefs about the population histories. In Figure 2 the red circles show the pup surveys and the red vertical lines show their standard errors. The black lines do not represent possible trajectories (the red and green lines do that) but envelopes of trajectories. The following is the most important thing I am going to say. Concentrate not on the lines but on the spaces between them. There are 7 lines, making 6 intervals of equal probability 0.16. This is supposed to be the probability that the true value lies within the interval. (The thing that confidence intervals don't mean but people always want to interpret them as if they did.) If you are asked to decide which interval the true value lies in, you can do no better than rolling a die, with a 5-to-1 chance that you will be wrong. Prudent management will take that fact into account. Moreover there is one chance in 25 that the true value lies outside any of the intervals.



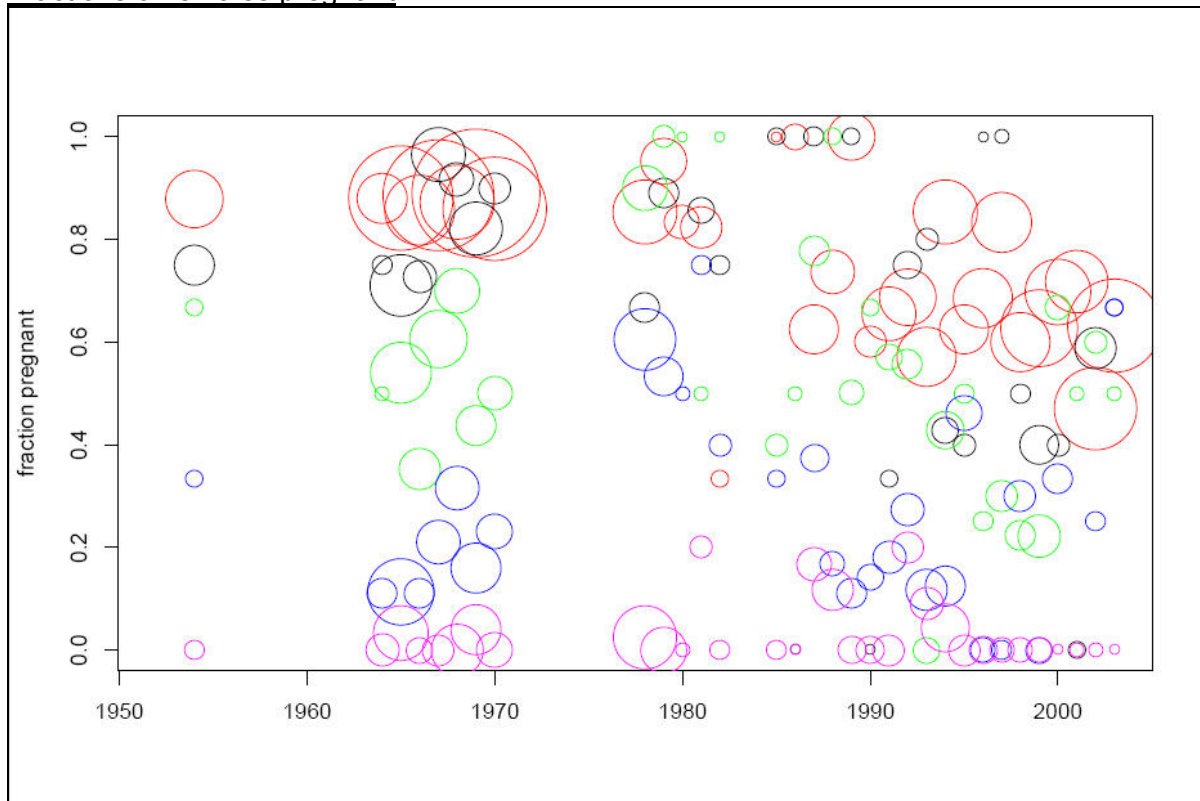
The small blue bars at each end are how the ends of the black lines change when the 2004 pup survey is taken into account. It reduces our uncertainty about 2004 pup numbers, and also about 1952 pup numbers.

Vague priors tend to be vague about things they notice, and dogmatic about things they don't. For example the model so far is dogmatic that the multiplier is constant over time. There are hints in the pup surveys that this may not be true, that the rate of increase has itself been increasing. Straight mathematics would strongly prefer the increasing-multiplier model. But we don't: it offends our prior beliefs about the population in the 1950s. 1956 had a pup catch of 360 thousand, the following year half of that. This casts doubt on the pup numbers back then, and also on the constant multiplier idea. Suggests there is scope for a better model but this is not it.

Preliminary work to get roughly independent auxiliary variables turns out to be crucial for getting the 3-parameter model to converge.

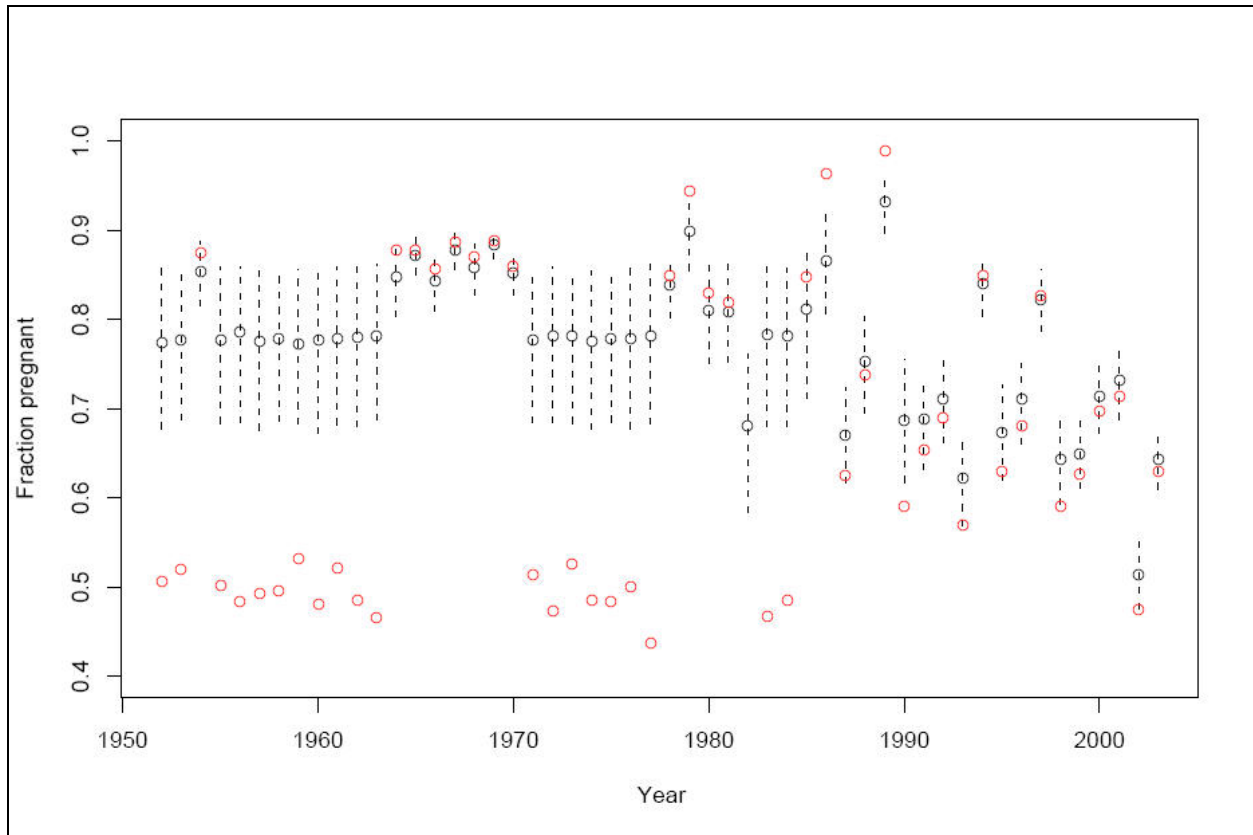
So far what we have done looks very little different from classical statistics, except that we end up making statements about the probability that a certain fact about the real world is true. It's almost "How to be Bayesian if you must." Though one benefit for all the extra work: we never assumed that all the measurement errors were equal, and in fact supplied a data set including the standard deviations for each survey. So we went beyond the simplest linear regression theory at no extra cost beyond the cost of being Bayesian.

Fractions of females pregnant

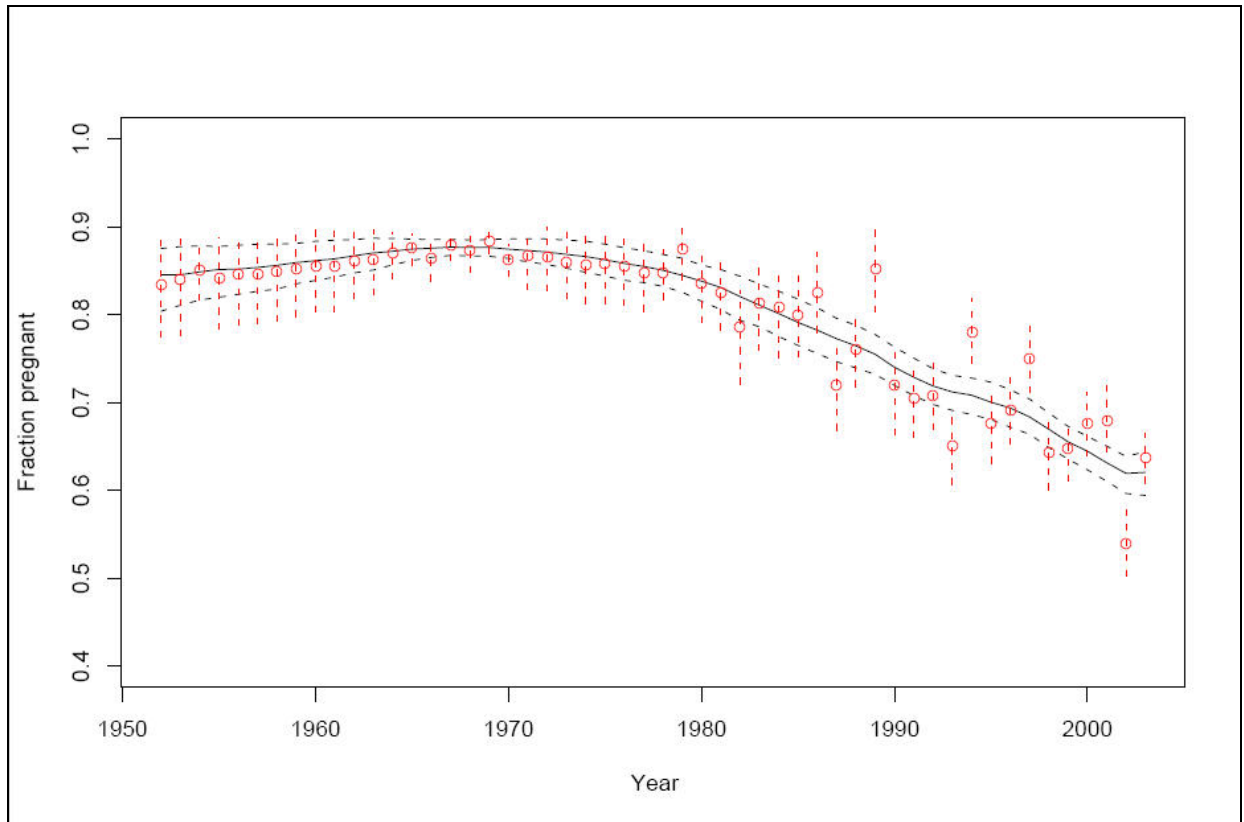


Now we get into some of the reasons why we might want to be Bayesian. One thing that might change with time is the fraction of females pregnant; and this is what we have the most data on. There is evidence of year-to-year changes and also of systematic changes; there are years of little or no data. We seek a description that will allow us to use a year's observations when they are abundant, and otherwise use observations of the same seals, in the same bit of ocean, in nearby years. Figure 3 shows fractions pregnant for females from age 4 (magenta; at time of giving birth; 3 at time of observation) to 8+ (red). The area of a circle is proportional to the number of females examined. This leads to a hierarchical model, when the true fraction pregnant in a year is a random variable drawn from some hyperdistribution, and then the observed number pregnant in a sample is a binomial random variable with the given true probability. We look in detail at age 8+ females.

We can reject immediately the hypothesis that the true fraction pregnant is the same in every year. The red symbols in Figure 4 are for a model with a vague prior distribution of mean 0.5 for each year separately. The black symbols are for a hierarchical model; the dashed vertical lines show the interquartile range of the posterior distribution.



Assuming that no year has information about any other year has, naturally, the best fit to data (lowest deviance), but has the largest number of parameters to be estimated and contains no information about unobserved years. How do we decide whether to prefer a model that fits better but has more parameters? Deviance Information Criterion (DIC), based on the estimated number of model parameters which is the difference between the average deviance and the deviance of the average prediction. (Deviance is a funny term: perfect fit has nonzero deviance.) I can't motivate it. It has been shown to work well in some situations and not in others. It tends to be used because we need to believe in something. A pure hierarchical model has a better DIC, but for unobserved years it pays no attention to evidence of a trend over time.



We have no parametric model for what a trend might be, and so in Figure 5 we use a moving average model for the hyperparameters alpha and beta of the Beta hyperdistribution. A pure MA model (black lines; the dashed lines are the interquartile range; similar to what Brian Healey used but with a different smoothing procedure) has worse DIC than pure hierarchical. A hierarchical MA model (red symbols), where the MA procedure is used to get the hyperparameters, has by a small margin the best DIC. This is not in itself a reason to select it, but it also accords better with our belief that there is a trend and (especially for making projections of future pregnancies) we will get more accurate answers by taking it into account. It has the appealing property of relaxing to the overall average of all the data many years away from all observed years.

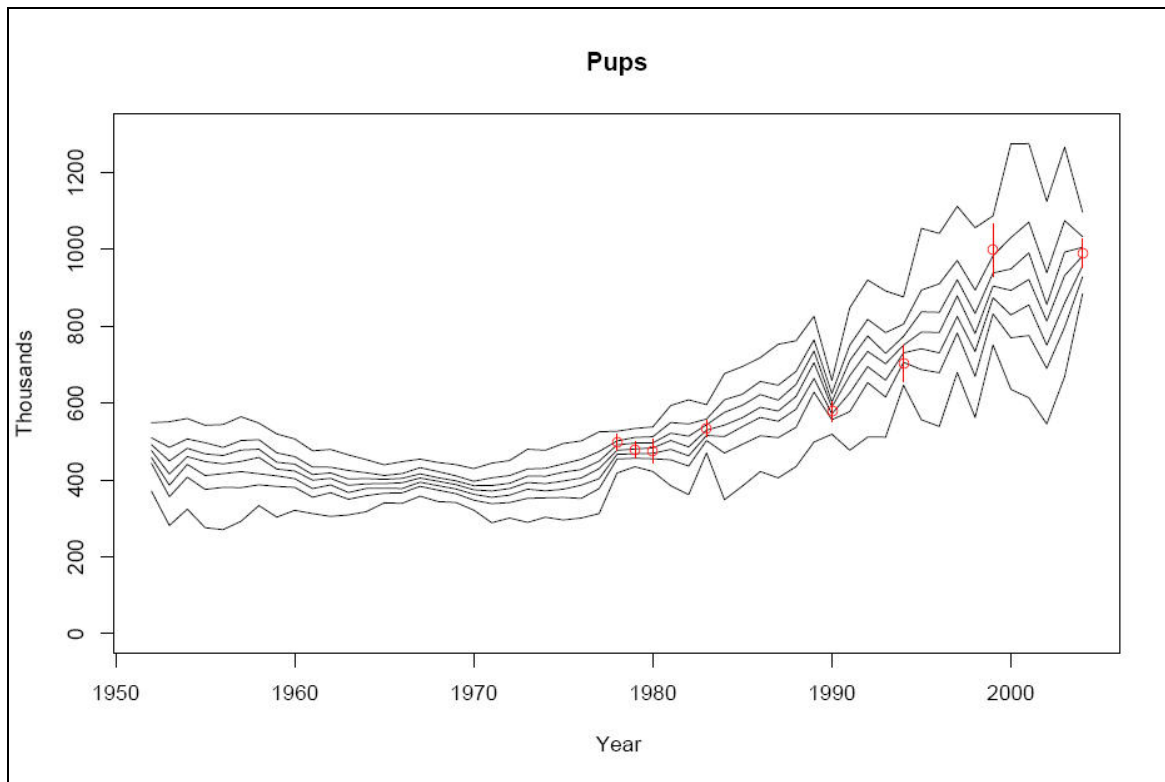
Full ballistic population model

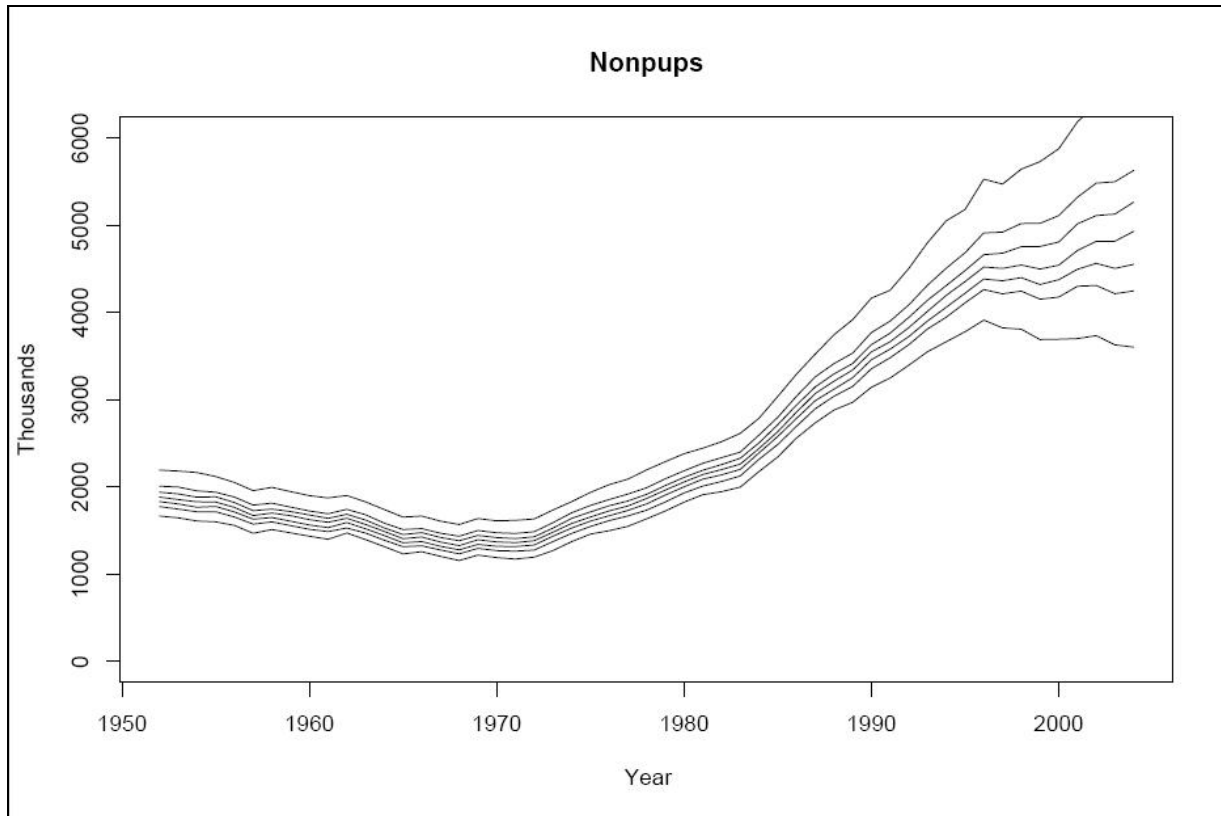
The term ballistic is used to indicate that each cohort is launched with its own pup numbers and then proceeds deterministically with the given natural mortality rate and catch removals. What new issues arise in the full model?

1) There are several ages each with its own fraction pregnant. But all ages give birth to identical pups and therefore, because we don't actually determine the age of a seal giving birth, there is no way to recover this information later. There could be 5000 more pups born to age 6 mothers and 5000 less to age 7, and we would never know. Thus the pregnancy fractions of separate ages are irredeemably confounded and we might as well estimate only the age 8+ fraction and adjust the others to have the same number of standard deviations from their expected values. Pragmatically, this also hugely reduces the number of parameters to be estimated and makes the program run much faster.

2) Pups have a higher natural mortality rate than adults. As a first step we assume pup natural survival is the square of adult survival over a year. In principle the power could be estimated, in practice it is confounded and in any case makes no different predictions about things we care about like seal numbers now.

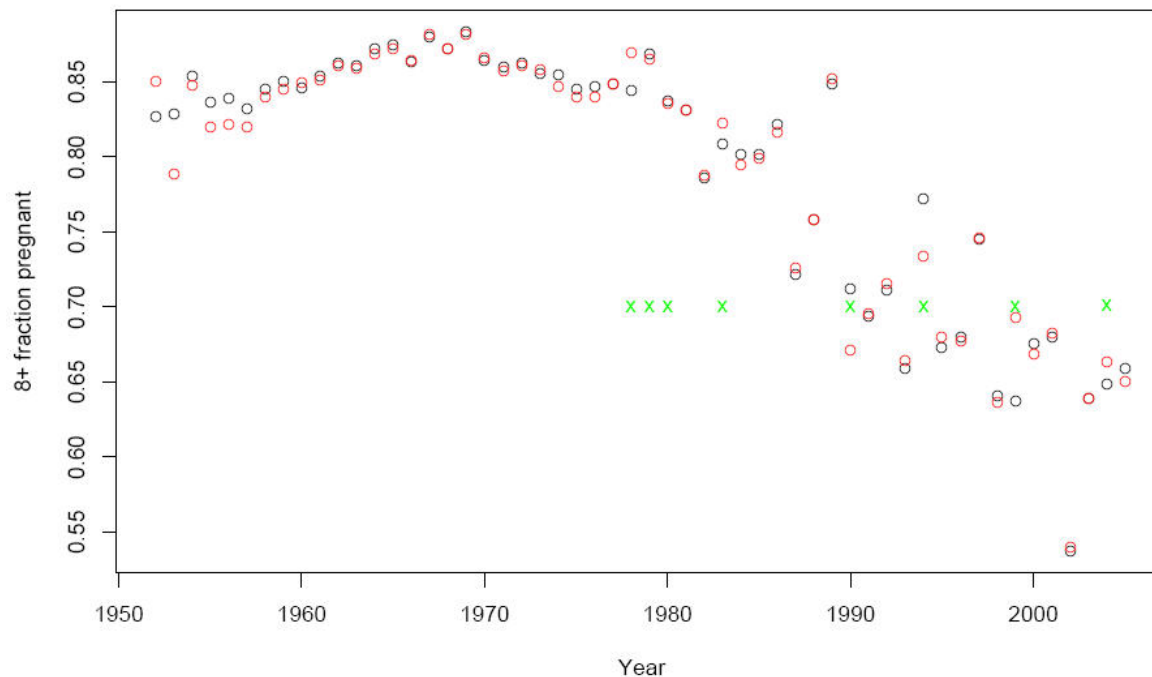
3) Initial conditions now are a vector of 8 nonpup numbers, not a single starting number (pups are computed from pregnancy fractions.) In the absence of nonpup surveys at any time (age distribution of catch is the closest we have) it makes sense to require that the starting age distribution is close to a stable age structure, including pups, and close to equilibrium. But we can't require that it be exactly that because we know the pup catch varies hugely between years and this makes the subsequent age structure far from stable. So, technically, we have another prior distribution for parameter values that make the starting population not too far from stable. It is a very indirect, implicit prior, which needs some tricks to define it satisfactorily and then specify it in JAGS. Figures 6 and 7 show the envelopes for pups and nonpups.





4) There are new questions we can ask, like the probability that the seal population has been declining between 1999 and 2004. The ratio of the two numbers is just another function of parameters whose marginal distribution we can sample.

5) One can also ask about the posterior distribution for fraction pregnant and whether it differs from the prior. It turns out it can in years when there are pup surveys (marked in green in Figure 8), and also at the beginning where stability was not defined very nicely. Some badly behaved exploratory runs converged to totally unbelievable population histories, and then compensated by choosing wild outliers for the fraction pregnant in survey years.



Deviance calculations estimate that there are 2.6 model parameters, barely more than the 2 of the model with exponential growth of pups. This means that, despite the posterior adjustments of pregnancy fractions in the years of pup surveys, there are few effective parameters associated with the freedom to choose pregnancy.

Discussion

Alf: how did you decide on the smoothing window?

Geoff: after a bit of play, it was fixed at 10 years.

Alf: how can you do a GOF test to the model

George: assess the predictive capacity at the higher level of hierarchy. DIC at the lower level of hierarchy. It is like cross-validation. Eg: new recruits per streams > low level; regional recruits is high level

Pierre: how does one project forward?

Geoff: With a pure hierarchical model, as soon as run out of data, you are back to projecting the average. With a running average, projections would be tied to the recent history. Now the full seal production model: age-structured model is simplified by assumed that pregnancies at all ages are perfectly correlated. May not be true but will not know anyway. Survival rates: strong belief that pups have low survival more than older ones but cannot get at it. Used informative priors in pregnancy rates. Also used informative priors to start off with a stable age distribution. Strong influence of data on posterior. Nonpup numbers smoothed because (missed the point made CHECK). Funny that precision greater than latest data close to latest survey

George: This problem arises in all sorts of age-structured models

Alf: it has to do with basic model. In some cases, will give exponential growth.

Kent: are filling pregnancy rates from priors when no data, which are narrower.

Steve: Taylor's law higher variance with higher mean

Geoff: Comment on Bayesian modelling: advantage: having gone to all this trouble to build this posterior, we also have random sample of all sorts of other parameters!!! Work real hard at first but have a lot of bonuses. Fig 8 of 8: black is prior median, red is posterior median; big differences where there is a pup count. We could argue that pregnant does not result in pup

Rob: why are results inconsistent in differences between red and black; why is different.

Mike H: Is this backward extrapolation?

Geoff: No, it is not a Peter Shelton model. Where go next: should use the age distribution of the catch. Let's you get at things like stable age distribution. We could get more out of it than are now.

Pierre: What software was used?

Geoff: I first used WinBUGS, then switched to JAGS (which runs in Linux) and stayed with it because it runs models where WinBUGS froze. Also, the author of JAGS is very good to respond to questions and make changes to his software. The syntax is a bit different from WinBUGS.

George: I really like the fact that JAGS has a way to simulate data and fit model to data; that is the best way to determine how much data is needed, i.e., to answer the question of data sufficiency. JAGS does not allow you to do cyclic walks, you can in WinBUGS. All are useful in their own way (JAGS, WinBUGS, OpenBUGS) because they offer different options.

Rod: if you put same model in all three, do you get same answer?

George: that is why JAGS was invented, to look at this question.

Geoff: I had correlation issues. JAGS is faster. Also JAGS code has been fixed.

Alf: for convergence, best to let it converge a long time vs using many chains

Geoff: it is easy to see if there is a lack of convergence. JAGS produces only one chain, you can do two runs separately and compare them. I ran models several times and checked if the trace showed no trend and answers were qualitatively indistinguishable over a few runs. Demonstration that there is convergence is difficult; non-convergence is easy

George: diagnostic statistics exist but looking at traces is better. The characteristics of the wiggle trace that shows convergence is that it looks random, not trendy. MCMC is not supposed to be a random walk. You can also look at the posterior distribution as the MCMC goes on. You can see it smooth out. But it is feasible to have a multi-modal distribution. Can you get convergence if there is a multi-modal distribution

Geoff: You would have trouble since the MCMC will update one mode at a time.

George: CODA and BOA have frequentist takes on Bayesian analysis.

Alf: when I run non-Bayesian model by likelihood and mean square, they often don't converge and when I simulate I found I can get crazy values of these parameters. If I get crazy results, then the model must be wrong. Problem arises when poor data and many parameters to estimate

George: All models are bad. All make assumptions that are untenable to varying degrees because they simplify biological processes. That influences the capacity to fit the data. There is no single approach to get you around this (in either Bayesian or empirical methods).

Geoff: This is the question of which has priority the parameter or the data.

George: you can have 20 years of data but did no perturbation of the population. In that case, you are forced to use an informative prior.

Geoff: if you insist on getting a definitive answer.

Alf: one problem is that all these parameters are changing over time. We should spend time on making better models. I am afraid of jumping to Bayesian when have problems.

George: That is a totally fair comment. The Bayesian approach has a suite of problems like others. And some are similar.

B-2 Working model 2: A Working Model: The Stock-Production Model in use for Assessments of the West Greenland Stock of Northern Shrimp. Carsten Hvingel and Michael Kingsley (presenter), Grønland Natur Institut.

A model of the dynamics of the West Greenland stock of the Northern Shrimp *Pandalus borealis* has been built and taken into use for the assessment of the stock and the formulation of management advice.

The model is not age-structured, and production is related to stock size by a Schaefer (quadratic) form¹. Each year, catch and predation by cod are removed from the stock and production is added; a lognormal process error is also included in proceeding from the stock in one year to the stock in the next. The model uses 5 input data series: catches (1955–date), standardised CPUE (1976–date), a survey index of biomass (1988–date), cod (predator) biomass (1955–date), and predation-rate estimates (1989–1992).

Predation is related to the biomasses of predator and prey by a Holling type III function, depending on 2 parameters, with a lognormal error term. The 4-year series of predation-rate estimates contributes to the estimation of the parameters of the predation relationship, but the model is not dependent on these estimates. Both CPUE and survey are considered indices of stock size, each with its own scaling factor (catchability) and lognormal error. Reported catches are considered absolute in size and error-free, partly because the objective is to provide advice on the management of reported catches, and are the only data series that controls the scale of stock size.

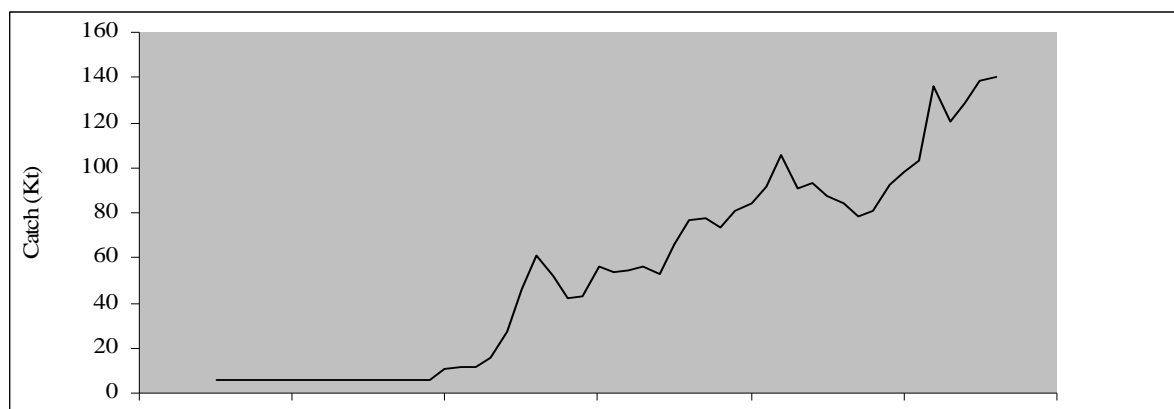
Because of the difficulty of getting absolute estimates of biomass, the working variable for stock size in the model is its ratio to the Maximum Sustainable Yield Level (Bmsy). Parameters affecting the process are therefore: error variances associated with the stock-dynamic process, CPUE, survey, predation in general, and predation from the 4-year series; catchabilities for the CPUE and the survey; 2 parameters defining the predation relationship; production rate at MSYL; and the MSY itself.

The model is built as a Bayesian model formulated for the WinBUGS platform. Non-informative priors are used, except for the maximum predation rate at high prey stock size (given an informative Normal prior with truncated tails based on experimental feeding studies) and carrying capacity (given a slightly informative log-Normal prior with infinite tails in order to smooth convergence), for both of which there is little information in the data. It runs smoothly and gives repeatable results, with significant updating of the priors. The modelled stock trajectory tracks both the biomass indices, and error variances are acceptably small.

¹ The model was originally formulated as a Pella-Tomlinson model, allowing a skewed relationship between production and stock size. The resulting stock-production curve was very close to Schaefer, and the simpler Schaefer model ran so much faster that the model was simplified to that form.

	Priors			Posteriors			
	Mean	SD	Median	Mean	SD	Median	Skewness
MSY	215	249.9	99.66	157.1	52.24	145.2	0.68
K	1300	1691	790.9	2181	3310	1515	0.60
Survey.catch	8.654	18.88	0.3153	0.41	0.1894	0.3978	0.19
CPU.catch	72.48	175.8	0.9993	1.061	0.4105	0.9781	0.61
P.50	4.346	11.85	0.3965	5.651	4.047	4.813	0.62
Omax	3	0.3166	3	2.999	0.3122	2.999	0.00
cv.CPUE	2378	283100	0.1518	0.03846	0.01484	0.03582	0.53
cv.Survey	96370	4.56E+07	0.3731	0.1591	0.0288	0.1558	0.34
cv.Process	14370	3.83E+06	0.2631	0.1014	0.02621	0.1021	-0.08
cv.Predation	567400	2.03E+08	0.4787	0.494	0.2943	0.4623	0.32
cv.Grunwald	203200	6.40E+07	3.354	0.8792	0.55	0.7623	0.64
P[1]	0.9703	0.1964	0.951	0.9172	0.1933	0.8931	0.37
Bmsy	650	845.7	395.5	1091	1655	757.4	0.60
Zmsy	0.8965	2.094	0.2521	0.1888	0.06094	0.19	-0.06

Table: Updating of priors in the W. Greenland assessment model for *Pandalus borealis*



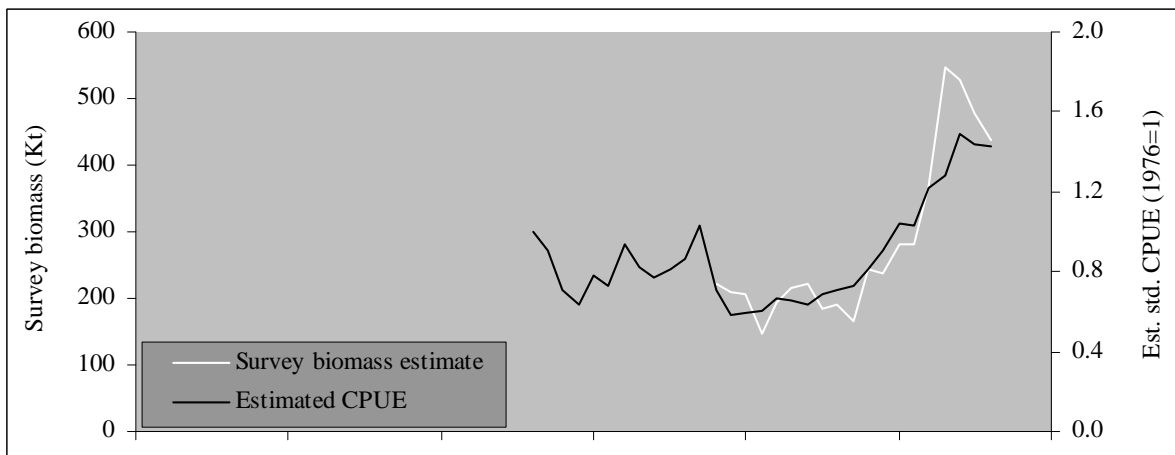
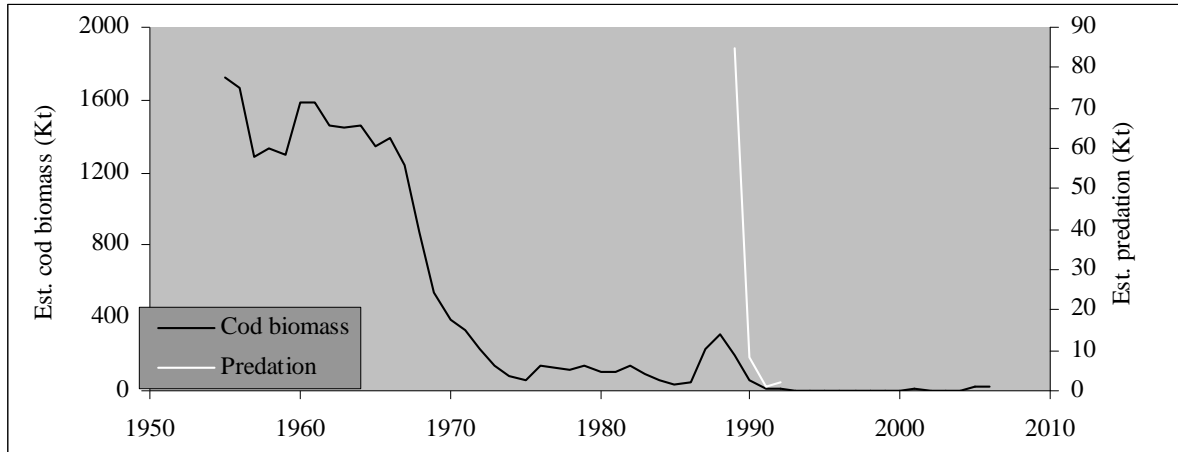


Figure. Data series input to the W. Greenland shrimp model: catches, survey and CPUE indices of biomass, and biomass of predator with a short series of estimated predation values.

The smooth working of the model is probably due to wide ranges in all the data series, good agreement between the two biomass index series, and a stock history that is consistent with all the data series: ‘the story hangs together’.

In formulating management advice, the model time series are extended into the future so that all the information on the joint distribution of the determining parameters is used in predicting the consequences of management actions. While other parameter estimates—such as error variances—are used in evaluating model fit, the principal stock-dynamic parameters that enter into formulating advice are the MSY and the most recent estimates of the relative biomass level. However, advice is most strongly affected by the predicted probabilities that given catch levels will take stock levels below, or total mortality above, precautionary limits. The model branches in the future so that it can simultaneously forecast outcomes for several catch levels, and several possible trajectories for the predator stock size, several years into the future.

Discussion

Alf: If you ask one expert to give a prior probability distribution, he will, and then you ask another expert and she will give another distribution. Probability is defined subjectively even if they both meant the same thing.

Michael K: One could ask people to give min, max and median and then one would fit a beta distribution. If you did that to many people, you could get at the prior distribution of the parameter.

Alf: if you get enough response, that becomes data that can be analysed; Bayes can disappear.

Michael K: That is a useful dataset for local knowledge of population qualitative statements.

Alf: It is wise to make subjective statements before you get out and get data. So it is a priori.

Michael K: This shrimp model actually works. Initially we used a Pella-Tomlinsen which was slow to run. We now use a Shaefer or logistic model. Relative estimate of biomass ($P = B/B_{msy}$) are also easier than estimating biomass directly. Catches were assumed error-free. CPUE and survey biomass were unscaled indices. Used WinBUGS. Zmsy was stock productivity at MSY. Omax was the max predation rate. We had convergence to the same results from different starting values. Priors are updated significantly by likelihood. Why did it work so well? The range of data on cod stock change is very large. The catch change 6 KT to 130 KT. Survey and CPUE also varied widely (by factor of 2). The story hangs together: there few shrimp when cod is high and shrimp increased when cod collapsed. The inter-quartile range of the posterior fit the survey data and CPUE quite well.

Pierre: The number of parameters estimated was greater than the question originally posed.

Michael K: Don't extract distribution parameters and then model them. Instead, simply extend the modelled period beyond the data. You get "free" estimates that way. This is not simply for efficiency. It is because you cannot easily extract the complexity of the posterior distributions and the joint multivariate distribution between variables. You push it for all its worth and get as much as you want. If you use distribution parameters and then approximate the distribution by a model, then you add uncertainty. Use a step function to calculate the risk probability.

Jack – How does one model for climate change?

Michael K– There is a big shopping list of things to model there. Global warming may increase cod.

George – estimating q for surveys is good – marine mammal surveys try hard to get it at one; have you tried estimating q . I treat my catches as priors. Marine mammal people don't have the catch data; inspection of fish catch is thorough.

Tim: shrimp loss rate is not recorded

George: what distribution was used for K ?

Michael: I used a lognormal; a uniform would cause WinBUGS to stop.

Some basics on frequentist-likelihood modelling

Let D be data (observations), θ a parameter (vector) and $f(D|\theta)$ the likelihood. In frequentist analysis the parameters are considered as constants, and it is only the data D that are treated as stochastic variables.

The maximum likelihood estimator, θ_{ML} , for θ , is defined as the value for θ that maximizes the likelihood, i.e., the value for θ that maximizes the probability of having observed the observations D . This is an appealing criterion, and it can be shown that θ_{ML} has nice asymptotic properties such as unbiasedness. In data poor situations, however, θ_{ML} may be rather biased.

To find the θ_{ML} in practice, it is often more convenient to maximize the log-likelihood, $L = \log(f(D|\theta))$. In complex models it is not a trivial task to find θ_{ML} , and it is wise to see how robust the maximizing algorithm is to the choice of initial parameter values.

Once the θ_{ML} is found, the properties of the estimator can be estimated by letting the θ_{ML} play the role as the true θ -value, and simulate data samples, D_{sim} , from the model $f(D_{sim}|\theta = \theta_{ML})$ with the same sample size as D . In each run θ_{Msim} is calculated based on D_{sim} , and statistics like bias, standard deviation, correlation among estimators, confidence intervals and confidence regions can be found. In addition one can calculate the same statistics on functions of the parameters, e.g. a 95% confidence interval for the predicted abundance the next year in a population dynamics model. In addition, model validation can be performed by comparing the properties of the residuals based on D with the properties of the residuals from the simulations. The success of a frequentist-likelihood approach depends strongly on an appropriate model and a sufficiently large sample of appropriate data. If this is not the case, θ_{ML} will often reveal this by providing “crazy” values, e.g. a negative instantaneous mortality rate.

Some basics on Bayesian modelling

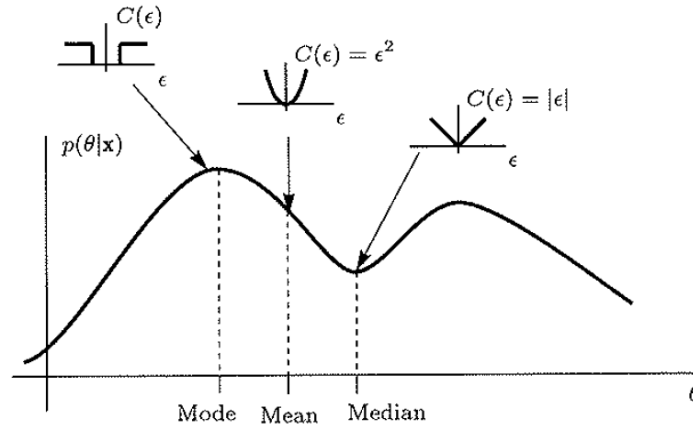
Let as before D denote data, θ a parameter and $f(D|\theta)$ the likelihood. In Bayesian analysis, however, the parameter θ is allowed to be stochastic. All inference about θ is based on the posterior distribution of the parameter conditional on data, $f(\theta|D)$:

$$(1) \quad f(\theta|D) = \frac{f(D|\theta) \cdot \pi(\theta)}{\int_{\theta} f(D|\theta) \cdot \pi(\theta) d\theta} = \frac{f(D|\theta) \cdot \pi(\theta)}{f(D)}$$

where $\pi(\theta)$ is the prior describing (subjective) a priori information of θ in terms of a probability distribution. Note that the denominator in eq.(1), $f(D)$, does not depend on θ and thus acts as a proportionality constant in the posterior. The calculation of $f(D)$ easily becomes prohibitive in complex models with many parameters.

The Bayesian approach has an intuitive appeal because it allows taking into account prior (expert) knowledge of a parameter, and it weights data versus subjective knowledge according to what source is most informative. Note however, that in contrast to the frequentistic approach, the probability concept no longer is defined precisely. Two experts with the same opinion may quantify their knowledge in terms of priors differently. In addition, a possible bias of a Bayesian estimator cannot be estimated.

To define a Bayesian point estimator, $\hat{\theta}_B$, the cost function $C(\theta)$ of the error $\theta = \hat{\theta}_B - \theta$ must be defined. $\hat{\theta}_B$ is then the value of θ that minimizes the Bayes risk $E(C)$ with regard to the posterior. For a square cost function, $\hat{\theta}_B$ is the mean of the posterior, while $\hat{\theta}_B$ is equal to the median for $C(\theta) = |\theta|$. For a uniform cost function $\hat{\theta}_B$ is the mode in the posterior. In this case $\hat{\theta}_B$ is found by maximizing the numerator in eq.(1), which technically is identical to finding a ML estimator.



(a) General posterior PDF

The last decades the application of Bayesian modelling has been increasingly popular due to the introduction of Markov Chain Monte Carlo (MCMC) simulations, which circumvents the problem of calculating the difficult denominator, $f(D)$. MCMC gives a recipe of how one can simulate samples that converge in distribution to the posterior distribution $f(\theta|D)$ without the need to calculate $f(D)$. In these simulations the posterior distribution of any function of the parameters (e.g. a future abundance prediction) is obtained as well. And one may construct credibility regions from the estimated posterior, which is the Bayesian analogue to confidence intervals.

One fundamental tool in MCMC to produce Markov chains of samples from $f(\theta|D)$ is the Metropolis-Hastings algorithm:

$$(2) \quad \alpha(\theta_i, \theta_{i+1}) = \min \left(1, \frac{f(\theta_{i+1}|D) \cdot q(\theta_i|\theta_{i+1})}{f(\theta_i|D) \cdot q(\theta_{i+1}|\theta_i)} \right)$$

where α is the probability of accepting θ_{i+1} as a new value of θ in the chain, and θ_{i+1} is generated by the sampler, $q(\theta)$. that the common factor $f(D)$ for $f(\theta_{i+1}|D)$ in the numerator and $f(\theta_i|D)$ in the denominator disappear, so the ratio in the algorithm is easily calculated. Remarkably, the chain will then converge in distribution to $f(\theta|D)$ for any choice of the sampler $q(\theta)$. A great challenge, however, is to find a sampler with optimal convergence properties in terms of efficiency. When the distribution of any parameter conditional on all the others is known, the Gibbs sampler can be applied. In this case $\alpha = 1$, so the sampling score can be said to be 100%. In WinBugs the sampler is found automatically, and Gibbs sampler is applied when possible.

Besides finding an appropriate sampler, a major challenge in MCMC is to decide when the chain has converged. A complicating factor is that in general the samples in the chain are auto-correlated, even in the case when the sampler provide independent samples of the parameters.

One convergence diagnostics tool is to run several chains with different starting values to span the parameter domain. If the between-chain variance become sufficiently small compared to the within-chain variance, this is a good indication of convergence.

On the priors

Another fundamental challenge in Bayesian analysis is to assess the sensitivity of the results, e.g. the prediction of next year abundance in a population dynamics model, with regard to the choice of priors. This is particularly related to so-called non-informative priors for which there are no subjective knowledge. Intuitively one would think that it is just to choose a flat prior for θ . But since one does not have any information of θ , one has no information of e.g. $\log(\theta)$ either, so the prior for $\log(\theta)$ should be flat as well. This, however, is not consistent, a flat prior on θ is not consistent with a flat prior on $\log(\theta)$.

There are some general guidance rules, though, for how to choose a non-informative prior. For example, it is recommended to apply a flat prior for pure location parameters, like μ in the normal distribution $N(\mu, \sigma^2)$, and to apply a prior proportional to $1/\sigma$ for a pure scale parameter, σ . These recommendations are deduced from sound invariance principles. A problem is that the integral of these priors are infinite, so these priors become so-called improper. Another challenge in complex models is that it is not always trivial to identify if a parameter is a location or a scale parameter.

A very particular prior is the Jeffrey's prior, which is insensitive to parameter transformations in the sense that it has a unique expression in terms of the Fisher information matrix.

Assume a population dynamics model contains 10 non-informative priors where one is uncertain what the appropriate priors should be. Say that one is interested to examine how sensitive the prediction of next year's abundance is with regard to the choice of non-informative priors, and two different candidates for each parameter is chosen. This gives $2^{10} \approx 1000$ different combinations. If the model takes one hour to run, the sensitivity test takes about 40 days. Thus efficiency is an important task, where the application of e.g. automatic differentiation can be a very useful investment.

The harp seal population dynamics model

We apply a simplified two-age model to illustrate the challenges involved in a frequentist-likelihood approach as well as a Bayesian approach.

Data

Catch data from 1875, only Russian and Norwegian catches
Pup abundance estimates with cv's 1998, 2000, 2000, 2002, 2003
Pup abundance indices 1968, 1970, 1973, 1976, 1980, 1985, 1988, 1991
Reproductive data: 1288 females 1962-1993
Age distribution of whelping females: 373 (1980), 401 (1988)

Variables:

N_1, \dots, N_n = abundance of 1+ animals from years 1 to n
 N_{01}, \dots, N_{0n} = abundance of pups from years 1 to n

Parameters:

K = "carrying capacity", N_{1+} abundance in year 1
 M_0 = mortality rate for pups

M_0	= $\exp(\tilde{M}_0)$ = survival rate for pups
M	= mortality rate for 1+ animals
m	= $\exp(\tilde{M})$ = survival rate for 1+ animals
F	= birth rate
f	= female reproduction rate
p_1, \dots, p_A	= pregnancy rate at age 1 to A
A	= number of ages

Model:

$$N_1 = K$$

$$N0_1 = Km \cdot (f / 2)$$

$$N_i = (N_{i-1} - C_{i-1}) \cdot m + (N0_{i-1} - C0_{i-1}) \cdot m_0, \quad i = 2, \dots, n$$

$$N0_i = f / 2 \cdot (N_{i-1} - C_{i-1}) \cdot m, \quad i = 2, \dots, n$$

$$\alpha \equiv (f / 2)m_0 = (1 - m) / m \quad (\text{equilibrium})$$

$$f = (F / m) \cdot \left((1 - m) \sum_{a=1}^{A-1} p_a m^{a-1} + p_A m^{A-1} \right)$$

Note that, except for the equilibrium assumption at year 1, the model above is exact if we allow all parameters to vary in time. If we knew the values of all parameters for all years, the model would exactly fit the true abundance in time. At the same time, the number of parameters in this case would be much larger than the number of abundance variables, the latter being twice the number of time steps (years) because it is a two-age model. Thus an infinite set of parameter values would exactly fit the true abundances.

In practice we are left with a limited sample size of pup abundance estimates (5) within a narrow time window (1998-2003), 8 relative pup abundance estimates in the period 1969-1991, and a long time series of catches (from 1875). For the parameters there are very sparse prior knowledge other than the fact that no parameter can be negative, and that f and the p 's cannot exceed one.

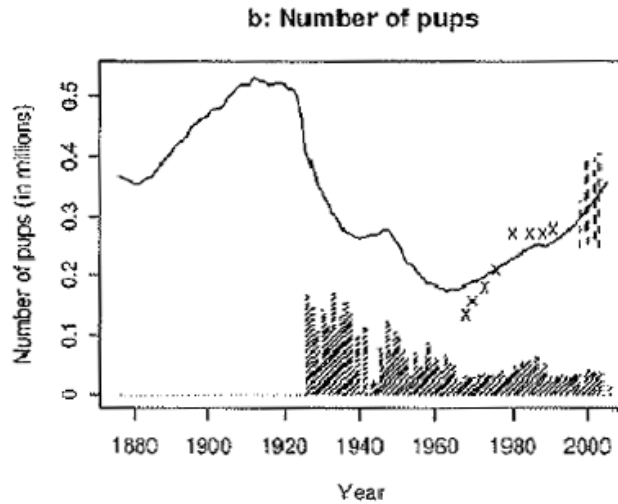
There are several ways to approach this typical data poor situation. My own philosophy is to start with as simplified models (few parameters) as possible to see how far these can be stretched, before more complex models are applied. Even simple models are hard to fully understand, a fact that in my opinion is far too often ignored.

In the model development I find a tight cooperation between the statistician and the biologist to be of uttermost importance. The Canadian sea mammal scientist Gary Stenson expressed the importance of such a cooperation very good during a visit to Tromsø recently, something like: *"The real potential of scientific progress evolves when the statistician becomes curious about biology and the biologist becomes curious about statistics in their cooperation"*

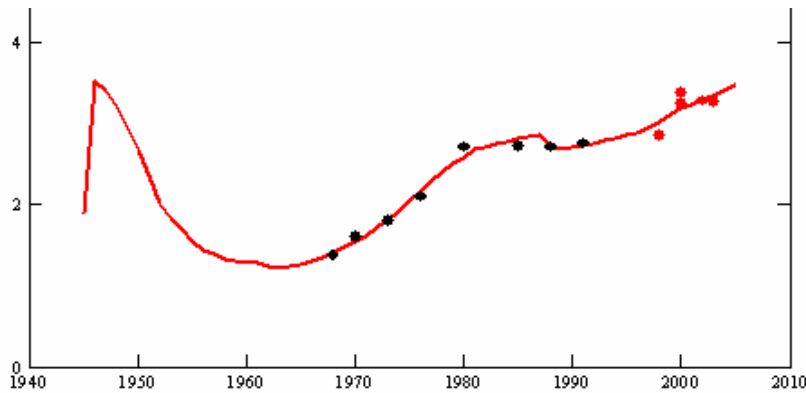
For the harp seal population it is a common opinion that parameters like instantaneous mortality rates may depend on abundance densities as well as on environmental factors. As a simple example I modelled the influence of density on M_0 as follows:

$$M_{0,t} = M_{00} \cdot \left(\frac{N_{1+,t-1}}{K} \right)^3$$

In this model the natural pup mortality increases rather dramatically as the 1+ abundance approaches the carrying capacity, K . An interesting feature of this model is that it provided much better fit to the absolute pup estimates and the pup index data than other (Bayesian) multi-age models with more parameters where the density regulation was put on other parameters, as shown in the figure below. This does not mean that one model is better than the other, but it illustrates the flexibility inherent in even simple models.



Bayesian model with more parameters than in right figure



Model with 4 parameters, multiply vertical scale with 100000.

Without the density regulation I have big problems in estimating $\hat{\mu}_{ML}$, e.g. when I try to estimate the 3 parameters K , M_0 , and f , and in some runs that work I may get non-biological results. When the same exercise is done applying MCMC with “biological” priors, apparently appropriate posteriors are produced. In my opinion this illustrates that one should really be cautious in applying Bayesian modelling in this situation: The MCMC technology works as a black box producing apparently reasonable results due to the priors, which easily hide the fact that the data model (the likelihood) in the first place is odd. Another experience with the frequentist-

likelihood approach is that several of the parameters covariate very strongly, which is an important information that will be obscured by applying independent priors.

Is the Bayesian approach worthwhile?

Pros: Bayes in general is particularly useful when there exist prior knowledge of parameters that can not be estimated from data. MCMC offers large flexibility and enables complicated models with many parameters. MCMC is an applicable tool that can be used by non-statisticians

Contras:

In data poor situations with problems of getting appropriate results by non-Bayesian approaches due to e.g. inappropriate models, apparently reasonable results can be obtained by Bayesian analysis, obscuring/camouflaging model failure. In data poor situations an important task is to apply a range of different models, reflecting different biological aspects that the biologist finds appropriate. A tool like MCMC is technically complicated in terms of e.g. interpretation, and might easily turn the focus too much away from the biology.

Recommendations:

Much effort should be put in developing appropriate data models in close cooperation between the statistician and the biologist, beginning with simple few-parameter models and a frequentist-likelihood approach to learn properly the properties of the model. Sensitivity analysis of priors should be emphasized in Bayesian analysis. Subjective knowledge should be quantified, experimentally when possible.

Discussion

- Is Bayesian probability a “vague” term for non-statisticians? Should we be doing Bayesian analyses because it is the “right” way, or because it is the easiest way? Different approaches to a problem will yield a better understanding of the underlying data and parameter relationships and values.
- The point is to get the biologists curious about statistics and the statisticians curious about biology (Stenson quote from Tromsø).
- Alf recommends a variety of approaches be applied to a problem such as traditional statistics, Bayesian models, sensitivity analyses etc.
- Discussion of AD model builder - fast in operation. ADMIT module in MATLAB may be another package of use.
- WinBUGS has an extensive help and example system in the main menu.
- Should we discuss finding a standard approach to Bayesian analyses (approaches and software choices)? E.g., JAGS may compile faster, and allow more complex models, than WinBUGS.
- OpenBUGS seems to give better feedback on model errors than WinBUGS, although either programme would be a good choice for beginning modellers.
- Great care must be taken in parameterization as the different programmes require parameters input in different ways.
- Data poor systems (e.g., bowhead pre-whaling population size) may constrain the functionality of the Bayesian approach (although this may be true of all statistical

approaches). The context of the problem will have a big impact on the efficacy of the various statistical approaches.

- SIRS versus MCMC approaches were discussed – some of the decision will be predicated upon the programming skills of the user. SIRS may be an easier programme to work with, if you are already a competent programmer; may be easier to understand the SIRS system for a simple model with relatively few parameters. Hobbs will prepare a description of SIRS to present.

B-4 Working model 4: Population dynamics of (mostly) short-beaked common dolphins from (at least) three stocks in the eastern Pacific Ocean.

George M. Watters, NMFS SWFC.

Using a model that is currently under construction, I provide working examples of various issues that arise in using the Bayesian approach to model the population dynamics of marine mammals, and although the title of my presentation gives emphasis to short-beaked common dolphins, my presentation is intended to be less about dolphins and more about modeling issues. I attempt to provide examples that illustrate the modeling process, and, since the work is still in progress, it is important that results illustrated in my presentation are not interpreted as final results. I am happy to discuss Bayesian modeling, either in general or for specific applications, with anyone. I have made my code available to participants of the Workshop, but, since this is an active area of my own research, I ask that users be discreet with its use and distribution.

I describe the dynamics of short-beaked common dolphins (*Delphinus delphis*) with an “extended” Pella-Tomlinson model that is cast in a state-space framework. These extensions include:

1. simultaneous fits to estimates of abundance developed from line-transect surveys in two regions and for three stocks (the northern, central, and southern stocks),
2. “movement” parameters which determine the probabilities that animals from the northern and southern stocks are observed in surveys that do not cover the entire range of these stocks (but one survey does cover the range of the central stock), and
3. time-varying carrying capacities for each stock.

I explicitly consider how environmental conditions affect movement by the northern and southern stocks and the carrying capacities of all three stocks. I also use auxiliary information collected from dip nets to index carrying capacity for the central stock and fit to a time series of such indices. I include both observation and process error in the model (this is essentially the state-space framework). Observation error is associated with the line-transect estimates of abundance, the environmental indices used to model time-varying carrying capacities, and the dip net data. Process error is included both in the sub-model that describes time-varying carrying capacity and in the overall dynamics equation described by the Pella-Tomlinson model. Posterior distributions for the parameters and other, derived variables of interest are estimated using the MCMC methods implemented in OpenBUGS, and the graphics provided in my presentation typify the graphics that can be produced from the OpenBUGS GUI (graphical user interface).

The remainder of this abstract is organized in subsections that address each of the issues which the Workshop Convenors identified as topics of specific interest.

Choice of prior distributions

My example model has numerous parameters, and I use a variety of approaches to specify their prior distributions.

1. I develop some priors on the basis of information in the literature. For example, results from Reilly and Barlow (1986) suggest that it is unlikely, but nevertheless possible, that the maximum rate of population increase at low abundance (r) is about 8% per year. Thus, I use a “folded normal” prior (by taking the absolute value of a normally distributed random variable) for

r and specify a variance for this distribution that allows one to infer that a value of 8% per year is not very likely.

2. I develop some priors by using approximations that provide a desired prior on a derived or implied parameter. For example, it might be desirable to place a uniform prior distribution on the maximum net productivity level (MNPL, expressed as the ratio of abundance to carrying capacity). However, in the familiar Pella-Tomlinson framework, there is not an analytical solution for z (the productivity shape parameter) in terms of MNPL; thus I cannot sample directly from a prior for MNPL and use algebra to convert these sample values into values of z . There is, however, an analytical solution for MNPL in terms of z , so I specify a prior distribution for z that provides an approximately uniform prior on MNPL over the range [0.5, 0.8]. Such an approximation is provided in my code.

3. I develop some priors by specifying means, variances, and ranges that provide biologically sensible results or do not cause numerical problems. For example, in a case where I compute the natural logarithm of the difference of two parameters with independent marginal prior distributions, I ensure that each marginal prior is specified in a way that does not allow the difference to be negative or extremely close to zero, either of which would cause numerical errors in calculating the logarithm. This type of problem can also be addressed by specifying appropriate joint priors, but I find that it can be difficult to communicate about joint prior distributions to some audiences (and even to myself!).

4. I develop some priors by adopting a suite of “common sense” rules. For example, I illustrate an S-shaped model that relates the “true” carrying capacities (K) estimated in the Pella-Tomlinson framework to “observations” of carrying capacity recorded in the dip net data. This S-shaped model has a parameter that defines an intercept, and I specify that the prior distribution for this parameter should include the possibility of an intercept near zero because it seems reasonable that dip net catches might be negligible when the true K is small (of course this does not have to be the case if there is reason to suspect a positive bias in observations of K).

5. I develop some priors on the basis of my own “prior belief.” For example, I use penalized splines to smooth some environmental data within the assessment model (illustrating how uncertainty associated with constructing environmental indices can be carried all the way through an assessment without having to do the smoothing “outside” the model), and I specify a prior belief about the degree of smoothness for these splines. I implement this belief by specifying a prior distribution which puts substantial weight on the hypothesis that one source of variance in the smooth (the variance of the random errors) is about twice as big as another source of variance (the variance of the smoothing parameters themselves). In another example, I control the degree to which carrying capacities can vary over time by using a box constraint to specify the prior belief that K cannot change (either positively or negatively) by more than 10% per year.

Convergence of MCMC chains and autocorrelation of posterior samples

In my opinion, convergence to a stationary distribution and autocorrelation in the posterior samples can often be assessed visually. I usually assess convergence to a stationary distribution by plotting “parameter traces.” I consider whether traces of single chains have properties like those which represent parameters that are sampled from prior distributions but are not modified by likelihoods. The key here is to recognize that the job is to assess convergence to a distribution rather than to a point. Therefore, do not expect parameter traces

to converge to straight lines. I also start multiple chains from different locations in the parameter space and consider whether these chains converge to similar distributions. I learned to make visual assessments of parameter convergence both by “training” my eyes in exercises where I viewed parameter traces in tandem with the results of more formalized convergence diagnostics (e.g., the diagnostics available in the R packages BOA and CODA) and by simply plotting traces of samples drawn from known distributions. Despite my reliance on visual diagnostics, formal tests may be a necessary standard in actual stock-assessment situations. It is important to recognize that an apparent lack of convergence (indicated either visually or by more formal tests) may actually be a result. For example, if the posterior distribution of a parameter is bimodal, the trace of that parameter may bounce back and forth between two separate regions of the parameter space. More generally, an apparent lack of convergence may usefully indicate that the estimation problem has multiple solutions. I evaluate autocorrelation in posterior samples using barcharts that plot the value of the autocorrelation coefficient (the heights of the bars) versus a lag that describes the distance between two samples from the same MCMC chain. Generally, I hope to see that the heights of the bars decrease rapidly with increasing lags and that this decrease bottoms out with bar heights that are close to zero (indicating little autocorrelation) at relatively small lags.

In my experience, difficulties with convergence and autocorrelation are frequent (with this frequency increasing as models become more complex) but can often be overcome with a few practical solutions. How a model is parameterized plays a major role in determining both the rate at which posteriors converge to stationary distributions and the degree to which posterior samples are autocorrelated. When I am confronted with slow convergence and autocorrelation, I usually consider how I might reparameterize my model. Centering and standardizing covariates (e.g., as one might do in the context of a simple linear regression) is often helpful. Convergence can also be slowed by prior distributions that are overly uninformative, and, therefore, when I am confronted with slow convergence I often re-evaluate my relevant prior distributions. If computing time is not important, a brute-force solution to autocorrelation is simply to produce huge posterior samples (e.g., millions or tens of millions) but then discard most of these results and keep, say, every 100th value. This process is called “thinning.” Quick checks of autocorrelation plots like those described above can be used to guide the selection of a useful thinning interval – i.e., try to pick a thinning interval which corresponds to a lag that has low autocorrelation for multiple parameters. There are formal diagnostics that can also be used to help identify a useful thinning interval. The drawback of thinning is, again, that many more samples are collected than are actually used in making posterior inference (thus it can be inefficient from a computing standpoint).

Sufficiency of data

In my opinion, the easiest method to assess whether data are sufficient for parameter estimation is to conduct a Bayesian analysis and compare the posterior distributions to their respective prior distributions, but the “best” method is probably to simulate data with specific properties (e.g., sample sizes, CVs, relationships to covariates, etc.) and determine whether these types of data are in fact sufficient. The easy approach is easy because when posteriors are different than their priors the data have obviously been sufficiently informative. Comparing posteriors and priors should not, however, be limited to comparisons of means; it is important to compare the distributions themselves (e.g., by comparing a range of quantiles, or making qq-plots). The more difficult, but preferable, approach may require a substantive investment in programming and subsequent analysis, but the payoff is gaining an understanding of data sufficiency more generally, with insight as to what type and quantity of data are “needed,” rather than knowing whether the data at hand are good enough.

Although the simulation approach mentioned above can help indicate the type and quantity of data that may be sufficient in the future, it does not alleviate the problem of having to deal with data that are not sufficient for parameter estimation in the present. There are various options for dealing with this situation (i.e., needing to develop a model, estimate parameters, and provide advice despite insufficient data), and I attempt to illustrate them in my presentation. One option is to be forthright about the lack of sufficient data and live with the uncertainty it brings. For example, the data illustrated in my presentation are not sufficient to estimate the MNPL (my posterior and prior for this derived parameter are the same), but I am comfortable with this result precisely because I am uncertain about MNPL and its value plays a critical role both in determining how a population responds to fishing and, at least in the context of the U.S. Marine Mammal Protection Act, in assessing the status of marine mammal populations. Another option is to bring an alternative source of data to bear on the same problem. For example, in an attempt to estimate how the carrying capacity of the central stock has changed I use both a series of sea-surface height data (which are the only type of data available for the northern and southern stocks) and observations of potential dolphin prey organisms collected during research cruises (which were not available for the other two stocks). Finally, it can be useful to conduct a sort of Bayesian sensitivity analysis where, in the absence of sufficiently informative data, informative priors and fixed parameter values are used to build contrasting scenarios. For example, in my model for *D. delphis*, I show how estimates of stock status (based on the ratio of abundance at any point in time to the abundance when a stock is at MNPL) are sensitive to different priors for parameters in the sub-model relating the “true” carrying capacity for the central stock to observations collected during research cruises as well as to different assumptions about the relative amounts of process and observation error in my model.

Software and programming issues

My presentation here is a simple set of bullet points – it is a bit difficult to combine these points into a coherent paragraph structure. Many of these points are specific to WinBUGS, OpenBUGS (which I use most frequently), and JAGS.

- I have found that programming skills in WinBUGS are transferable to OpenBUGS and JAGS (and vice versa). The syntax and programming flow in all three packages is similar but sufficiently different that code developed for one package will usually require some editing before it can be run in another package.
- It is important to understand how each modeling package parameterizes its distributions (e.g., by specifying estimates of variance or estimates of precision).
- MCMC chains can be thinned at two times during a modeling session within WinBUGS and OpenBUGS, during sampling or after sampling. Users should be aware that if thinning is conducted after sampling some of the default graphics will not be developed from the thinned chain. To ensure that graphics are developed from thinned chains it seems best to thin during the sampling process, this can also reduce memory requirements and the sizes of output files.
- I often aim to include various bits of “data processing” within my models (e.g., in my presentation I illustrate smoothing a series of environmental data within the larger model of dolphin population dynamics). The advantage of such a thing is that uncertainty associated with the data processing (e.g., how much to smooth the environmental series) can be integrated into the assessment and carried through into the results that are relevant to providing management advice. The disadvantages of this approach include added model complexity and increased computing time.

-
- Following from the point above, I also aim to predict as many “values of interest” (e.g., metrics like the ratio of abundance at any point in time to the abundance when a stock is at MNPL and estimates of sustainable mortality like potential biological removals, PBR) within my models as possible. The advantages here are that, again, uncertainty is integrated into these values of interest and their posterior distributions are computed as a matter of course, but this comes at the price of increased computing time.

Discussion

- Geoff: left out the key PRO thing is that calculates the things that conf limits say they do but they don't
- Alf: yes; but for me it does not solve anything because the Bayesian probability is vague since includes prior.
- Alf: the best thing is to do both empirical and Bayesian.
- There are insights to be gained from a progressing from deterministic, sensitivity analysis, to empirical, to Bayesian (also sensitivity) and insights are gained from all methods
- Geoff: ADmodel Builder does all those things at a price
- Matlab also had Admit module [ask Alf]
- Alf: Gaussian Markov Random Fields GMRF is another branch of stats that is very efficient. I have seen an example where they replace MCMC. Uses full conditionals
- Rod: Discuss standardization Bayesian methods and software
- George: not in favour; stifles creativity
- Alf: WinBUGS demo
- Alf: recommend WinBUGS but if make loops in WinBUGS do not repeat variable in model code
- George: any oddity in WinBUGS is when use lognormal you have to get the $\log(x)$ and put it in $\text{dlnorm}()$
- Michael: be careful at all distribution definitions; they are different from other software.
- George: I use R to look at distribution and then BUGS is totally different
- Pierre: sample it to see if correctly specified
- George: precision rather than SD
- Alf: Compiling is often problematic in WB
- Geoff: JAGS will compile models that WB does not
- George: cyclic graphs take long to compile
- OpenBUGS gives more information than WB but still unclear what say
- OpenBUGS also has diamonds that can click on and sometimes track through and figure out
- WB error messages – scour the WB listserv – few answers but often no answers
- As a tool for getting into it, WB is great
- Steve: Data poor systems are problematic ; maybe we chose a bad case to start
- Initial understanding is that data poor situations can be solved by MCMC but am thinking differently now.
- George: if have limited data, a lot of modelling approaches regardless, all have problems
- Michael: if try to do likelihood methods, you run into very slow things; don't know how to build into the likelihood method the process error. Easier in WB to build large models.
- I checked a model in Excel and got same
- The bonus is the extras parameters estimates
- George: is context dependent
- Huge models in ADModel builder take a lot more work than in WB. It is context-dependent.
- Consensus that WB is a good tool to beginners
- George: alternative for beginners is to do SIR if competent programmer.
- You can invest effort in learning WB or in programming a SIR also in language competent in

-
- SIR is 10 lines in Fortran
 - MatLab 10-15 lines
 - Why are people not using SIR?
 - SIR is more straightforward to understand if small vector size
 - No search routine involved
 - If posterior has oddities, like more than one maximum, you will find it easier by SIR rather than MCMC
 - Can someone prepare
 - Geoff: model says that if SSH above normal, K goes up forever
 - George: true, not totally satisfied by the formulation
 - A prior belief is that K increases slowly K_{t+1}/K does not change more than 10% per year. This debatable. May be something for sensitivity analysis.
 - Rob: what if there is a sudden change in current, that would be dampened by your
 - GIVES example of alternative parametrization for K that runs faster but worse mixing;
 - Smoothing depends on how much $1/sd^2$ is; how smooth or how rough do you want this to be
 - Prior weight is that there is more process error than observation error.
 - Can do all that in one model rather than separately
 - $g(K)$ there is information about inflection point
 - Sensitivity analysis
 - [QUESTION]: how to decide what constraints to place on error terms in priors]
 - Sensitivity shows bias-variance trade-off is gnarly.
 - RESULTS
 - WinBUGS graphics are given; not so bad
 - N/MNPL if below 1 depleted
 - N as fished / N if no-fishing shows that fishing does little to changes.
 - Can plug in uncertainty from Bayesian model into PBR calculation
 - Red line assumes MNPL at 0.5K; Black dots is using the MNPL in model estimates. Black dots are lower.
 - Did a second run with reparametrized parameters to do PBR

**B-5: Model 5: Bayesian population viability analysis (PVA) for the Cook Inlet, Alaska beluga population using a Sampling-Inference-Resampling (SIR) algorithm.
Rod Hobbs, NMML, NOAA/NMFS, USA**

A detailed population model was developed for the Cook Inlet beluga to assess the extinction risks faced by this small population. The model included immature and mature phases of both sexes (i.e., age- and sex-structured) and focused on the behaviour of a declining population at sizes less than 500 belugas. Small population effects were taken into account by examining survival and fecundity under a range of scenarios that considered demographic stochasticity, harvest, density dependence, Allee effects, constant mortality effects (e.g., predations), and unusual mortality events (e.g., catastrophes). Details of the model are available in Hobbs et al. 2006.

The PVA analysis was conducted to produce 3 results of interest:

- 1) A posterior distribution of the annual growth multiplier for the population.
- 2) A posterior distribution of population projections to 300 years.
- 3) A comparison among different types of mortality and fecundity effects in small populations.

A Sampling-Inference-Resampling (SIR) algorithm was used to generate a set of 10,000 values of the annual growth multiplier and associated parameters drawn from the posterior distribution. The annual growth multiplier is a derived parameter resulting from the choice of model parameters such as survival and fecundity probabilities. Although different survival and fecundity values result in the same annual growth multiplier each parameter set will have a different likelihood. For this analysis 100,000 trial parameter sets are drawn (Sampling) the model is run and the likelihood is calculated for each trial set (Inference) and a subset is drawn with replacement from the trial set weighted by the likelihood (Resampling) to develop the posterior parameter set. The advantage of this method being that this posterior parameter set then contained all of the necessary information to generate the products of interest without further sampling.

Data available to the analysis included 13 years of annual abundance estimates and harvest data for the Cook Inlet population. Life history and population parameters estimated for this and other beluga populations are available from the literature. Life history parameters of particular interest for modeling purposes were: survival probability, birth interval, age at first birth, gestation period, and lactation period.

The model was set up with nearly independent uniform priors for the population abundance in 1994, N_{1994} , and the annual growth multiplier, ϕ_0 . Other parameters for density dependent survival, and fecundity were drawn from priors informed by values from the literature. For computational convenience, if N_{1994} had a likelihood greater than $10^{-6} \times$ likelihood of $N_{1994} = \bar{N}_{1994}$ (the abundance point estimate for 1994), the parameters and age structure were retained as inputs to the Bayesian analysis (parameter values with lower likelihood would have no influence on the results). Each population was then projected from 1994 to 2005 and likelihood was calculated as:

$$L_j = \prod_{t=1994}^{2005} T\left(\frac{N_{t,j} - \bar{N}_t}{\bar{N}_t CV(\bar{N}_t)}, DF = 10\right) \quad (7)$$

where,

L_j is the relative likelihood of the j th population projection;

$T(X, DF = 10)$ is the density of Student's- t distribution at X with 10 degrees of freedom;

$N_{t,j}$ is the population size of the j th projection in year t , and

$\bar{N}_t, CV(\bar{N}_t)$ are the estimated abundance (point estimate) and associated coefficient of variation in year t .

Projections to 2005 with likelihoods less than $10^{-10} \times$ the maximum possible likelihood (i.e., the likelihood if the model N_t was equal to the abundance point estimate in all years) were discarded as having no contribution to the posterior distribution. A Sampling-Importance-Resampling (SIR) algorithm was followed in which the acceptable parameter sets were weighted by their relative likelihoods from projections to 2005, and a resample drawn with replacement to give a posterior distribution of outcomes. Projections to 2305 (300 years into the future) for this posterior parameter set were done to estimate the probability of decline and extinction during that period. Model comparisons between the various models were done using the Bayes factor, calculated as twice the natural logarithm of the ratio of the average likelihoods of the two resamples:

$$BayesFactor(x, y) = 2 \ln \left[\frac{\sum_{j=1}^{SIRtot} L_{j,x} / SIRtot}{\sum_{j=1}^{SIRtot} L_{j,y} / SIRtot} \right] \quad (8)$$

where $BayesFactor(x,y)$ is the Bayes factor comparing model x and model y ; $\ln[\]$ is the natural logarithm of the value in $[\]$, $L_{j,x}$ and $L_{j,y}$ are the likelihoods of the j th projection of model x and model y , respectively; and $SIRtot$ is the number of projections in the SIR subsample.

Where the Bayes factor had absolute value greater than 2 the model with the higher average likelihood was considered to be the more likely of the two, otherwise the models were of equivalent likelihood.

All models were compared to the Baseline model ($\phi_0 = U[0.94, 1.06]$) using the Bayes factor.

The three options for modifying the Baseline model, the constant mortality effect (C), the unusual mortality event (P_{Me}) and the Allee effect (A) were each considered. The time series of abundance and harvest data covered a sufficient range of population sizes (270-660 belugas) to compare between the Baseline and the Healthy Population models but not among the remaining options. Six models with the modifiers for survival and fecundity were considered, three of the Baseline with the C parameter at 1, 2 or 5 mortalities per year, one of the Baseline with the Allee parameter at 0.50, one of the Baseline with an unusual mortality event ($P_{Me} = 0.05$), and one of the Baseline with the C parameter at 1 and an unusual mortality event ($P_{Me} = 0.05$). Two additional models were included to test the sensitivity of the parameters: a Baseline and Healthy Population model that included an unusual mortality event and a C of five mortalities. For each model, 100,000 trials were projected to 2005 and the likelihood was calculated. Each population projection was fully defined by 13 parameters, however the parameter of interest was ϕ_0 , a derived parameter. A sample of 10,000 of these trials, weighted by the likelihoods, was drawn with replacement for the SIR algorithm resample for further analysis. For all populations the population size in 1994, 2005, 2105, 2205 and 2305 was retained and for declining populations the year that the population dropped below 200, 100, 10, and 2 animals was retained. A population with 1 or 0 individuals was considered extinct.

Reference

Hobbs, R.C. K.E.W. Sheldon, D.J. Vos, K.T. Goetz, and D.J. Rugh. 2006. Status review and extinction assessment of Cook Inlet belugas (*Delphinapterus leucas*). AFSC Processed Rep. 2006-6, 74 p. Alaska Fish. Sci. Cent., NOAA Natl Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Discussion

Time did not allow discussion of this model. Discussion of its methods was deferred to the general discussion on the following day. The results of that discussion are presented in the main body of the report.

Appendix IV

Additional references on Bayesian methods and modelling

Books:

- Albert, J. 2007. Bayesian Computation with R. Springer. x + 267 p.
- Bolker, B.M. 2008. Ecological Models and Data in R. Princeton University Press. vii + 396 p. *[Excellent text on ecological modelling methods, with many R code examples, including Bayesian modelling.]*
- Clark, J.S. 2007. Models for Ecological Data: An introduction. Princeton University Press. xiii + 617 p. *[Excellent text on ecological modelling methods, including Bayesian modelling.]*
- Clark, J.S. 2007. Statistical Computation for Environmental Sciences in R: Lab Manual for Models for Ecological Data. Princeton University Press. *[R code for modelling methods, including Bayesian modelling.]*
- Gelman, A., Carlin, J.B., Stern, H.S., and D.B. Rubin. 2004. Bayesian Data Analysis. Chapman and Hall /CRC Press. 668 p. [Advanced text on Bayesian analysis]**
- Hacking, I. 2001. Introduction to Probability and Inductive Logic. Cambridge University Press. 302 p. *[A careful discussion of the issues by a philosopher, with a lot of instructive puzzles (Geoff Evans)]*
- Hilborn, R. and M. Mangel. 1997. The Ecological Detective: confronting models with data. Princeton University Press, xvii + 315 p. *[Excellent introductory text on likelihood and Bayesian population modelling methods.]*
- Jaynes, E.T. 2003. Probability theory: the logic of science. Cambridge University Press, Cambridge, UK. 758 p. *[text on probability theory; a wonderfully polemic and dogmatic view (G. Evans)]*
- King, R., Morgan, B.J.T., Gimenez, O. and S.P. Brooks. 2010. Bayesian Analysis for Population Ecology. Chapman and Hall/CRC Press. xiii + 442 p. *[Excellent text on Bayesian ecological modelling methods, including population models, with appendices on R and WinBUGS Bayesian programming.]*
- McCarthy, M.A. 2007. Bayesian Methods for Ecology. Cambridge University Press. xiii + 296 p. *[Excellent text on Bayesian modelling methods, with a primer on WinBUGS use.]*
- Punt, A. and R. Hilborn. 2001. BAYES-SA - Bayesian Stock Assessment Methods in Fisheries - User's Manual, <http://www.fao.org/DOCREP/005/Y1958E/y1958e00.htm#Contents> *[Good primer on various Bayesian Fisheries population modelling methods; downloadable Excel spreadsheets are used to demonstrate the algorithms. Good reference list!]*

Papers:

(See also references listed above in Introduction)

Innes, S. and R.E.A. Stewart. 2002. Population size and yield of Baffin Bay white whale (*Delphinapterus leucas*) stocks. NAMMCO Scientific Publications 4: 225-238.

Wade, P.R., 2002. Bayesian population viability analysis. P. 213-238 In: Beissinger, S.R., McCullough, D.R. (Eds.), Population Viability Analysis. University of Chicago Press, Chicago. 577 p.

Appendix V

Software links

ADMB Foundation: <http://admb-foundation.org/>

ADMB Project: <http://admb-project.org/>

Bayesian Computation with R: <http://bayes.bgsu.edu/bcwr/>

Bayesian Methods for Ecology: <http://arcue.botany.unimelb.edu.au/bayes.html>

BOA: <http://www.public-health.uiowa.edu/boa>

CODA: <http://www-fis.iarc.fr/coda/>

JAGS: <http://www-ice.iarc.fr/~martyn/software/jags/>

MatLab: <http://www.mathworks.com/>

OpenBUGS: <http://mathstat.helsinki.fi/openbugs/>

PBSadmb (ADMB from R): <http://code.google.com/p/pbs-software/>

R: <http://cran.r-project.org/> and
<http://cran.r-project.org/web/views/Bayesian.html>

Running WinBugs and OpenBugs from R: <http://www.stat.columbia.edu/~gelman/bugsR/>

SIR method: No canned software is available but see Hobbs et al (2006), Hilborn and Mangel (1997) and Punt and Hilborn (2001) and Wade (1999) for examples, spreadsheets or pseudocode.

WinBUGS: <http://www.mrc-bsu.cam.ac.uk/bugs/>