NOAA Scientific Publications Report

JANUARY 4, 2013

HIGHLIGHTED ARTICLE

Declining abundance of beaked whales (*family Ziphiidae*) in the California Current large marine ecosystem.

Long-term population size of the North Atlantic humpback whale within the context of worldwide population.

Deep-sea Benthic Footprint of the *Deepwater Horizon* Blowout

ADDITIONAL ARTICLES

Atrazine fate and transport within the coastal zone in southeastern Puerto Rico.

Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management

Intracalibration of particle detectors on a three-axis stabilized geostationary platform

Have Aerosols Caused the Observed Atlantic Multidecadal Variability?

OTHER REPORTS, BOOK CHAPTERS, & INTERNAL PUBLICATIONS

Commercial Fishing in the Stellwagen Bank National Marine Sanctuary: Total Economic Value Baseline Assessment

Technical Appendix: Knowledge, Attitudes and Perceptions of Management Strategies and Regulations of the Gray's Reef National Marine Sanctuary by Users and Non-users of the Sanctuary

Mississippi River Freshwater Diversions in Southern Louisiana: Effects on Wetlands Vegetation, Soils, and Elevation

Statistical analyses to support guidelines for marine avian sampling: final report

HIGHLIGHTED ARTICLE

AUTHORS

Moore JE, Barlow JP (NOAA/SWFSC)

PUBLICATION DATE

On-line January 2013

TITLE

Declining abundance of beaked whales (family Ziphiidae) *in the California Current large marine ecosystem.*

JOURNAL

PLoS ONE

SIGNIFICANCE

Analysis shows strong evidence of decline abundance trends for beaked whales (Cuvier's beaked whale and *Mesoplodon* spp.) in the California Current from 1993 – 2008

Drivers of these apparent declines are unknown. Fisheries impacts can probably be ruled out. Impacts from anthropogenic noise and ecosystem change are plausible explanations, but additional research is required to evaluate these hypotheses.

Results may have implications for status determination (e.g., strategic, depleted) of beaked whales under the MMPA.

Beaked Whale

SUMMARY

Beaked whales are among the most diverse yet least understood groups of marine mammals. A diverse set of mostly anthropogenic threats necessitates improvement in our ability to assess population status for this cryptic group. The Southwest Fisheries Science Center conducted six ship line-transect cetacean abundance surveys in the California Current between 1991 and 2008. Sightings from these surveys were used to estimate abundance and model population trends of beaked whales. They also compiled records of beaked whale stranding events (3 genera, at least 8 species) on adjacent beaches from 1900 to 2012, to help assess population status in the northern part of the California Current. Researchers discovered strong evidence for declining beaked whale abundance in the study area. Cuvier's beaked whale (*Ziphius cavirostris*) had an average rate of decline of 2.9% per year and *Mesoplodon* spp. (pooled across species) had an average rate of decline of 7.0% per year. There was no evidence of abundance trend for Baird's beaked whale (*Berardius bairdii*). Stranding data were consistent with the survey results. Causes of apparent declines are unknown. Direct impacts of fisheries (bycatch) can be ruled out, but impacts of anthropogenic sound (e.g., naval active sonar) and ecosystem change are plausible hypotheses that merit investigation.

HIGHLIGHTED ARTICLE

AUTHORS

Kristen Ruegg, Howard C. Rosenbaum, **Eric C. Anderson (NOAA/SWFSC),** Marcia Engel, Anna Rothschild, Scott C. Baker, and Stephen R. Palumbi

PUBLICATION DATE

On-line December 2012

TITLE

Long-term population size of the North Atlantic humpback whale within the context of worldwide population.

JOURNAL

Conservation Genetics

SIGNIFICANCE

Controversy surrounds estimates of humpback whale populations "before whaling." The first genetic estimates had a large degree of uncertainty that was perhaps not well accounted for in the initial papers.

Considerable uncertainty remains; however this paper represents a fair and honest attempt to capture that uncertainty and include it in their estimate of long-term population size.

North American Humpback Whale

SUMMARY

Once hunted to the brink of extinction, humpback whales in the North Atlantic have recently been increasing in numbers. However, uncertain information on past abundance makes it difficult to assess the extent of the recovery in this species. While estimates of pre-exploitation abundance based upon catch data suggest the population might be approaching pre-whaling numbers, genetic diversity estimates based on mitochondrial DNA suggest they are still only a fraction of their past abundance levels. The difference between the two estimates could be accounted for by inaccuracies in the catch record, by uncertainties surrounding the genetic estimate, or by differences in the estimates timescales. The authors report an estimate of long-term population size based on nuclear gene diversity, increasing reliability of their estimate by increasing sampling across the geographic range and by analyzing more locations within the genome. They report an estimate of long-term population size in the North Atlantic humpback of ~112,000 individuals (2–3x higher than estimates based upon catch data). The remaining discrepancy between genetic and catch-record values highlights a need for continued evaluation of whale population growth and shifts over time, and continued caution about changing the conservation status of this population.

ATTACHMENT

PDF: See attached

HIGHLIGHTED ARTICLE

AUTHORS

Paul A. Montagna, Jeffrey G. Baguley, Cynthia Cooksey, Ian Hartwell, Larry J. Hyde, Jeffrey L. Hyland, Richard D. Kalke, Laura M. Kracker, Michael Reuscher, and Adelaide C.E. Rhodes

PUBLICATION DATE

TBD

TITLE

Deep-sea Benthic Footprint of the Deepwater Horizon Blowout

JOURNAL

Science

SIGNIFICANCE

The bottom biological communities surrounding the Deepwater Horizon oil spill were measurably damaged radiating out from the wellhead and along a plume toward the southeast.

Damage assessments for benthic impacts due to Deepwater Horizon must take into account recovery periods that may span decadal time frames.

Damaged communities were not the result of natural seeps.

www.darrp.noaa.gov/northeast/vieques/restore.html

SUMMARY

Oil from the Deepwater Horizon blowout reached the bottom of the Gulf of Mexico. Sediment cores were collected during Response cruises in fall 2010 to measure potential impacts on macrofauna and meiofauna – the two main soft-bottom benthic invertebrate groups. Changes in abundance and diversity of these fauna extend to 3 km from the wellhead in all directions covering an area about 24 km2 (and up to 17 km towards the southwest and 8.5 km towards the northeast, covering an area 167 km2). Benthic effects were correlated to total petroleum hydrocarbon, and distance to the wellhead; but not distance to hydrocarbon seeps. Recovery rates in the deep sea are likely to be slow, on the order of decades or longer.

ADDITIONAL ARTICLE

AUTHORS

Thomas L. Potter, David D. Bosch, Angel Dieppa (OCRM-NERR), David R. Whitall (NOS-NCCOS), Timothy C. Strickland

PUBLICATION DATE

~January, 2013

TITLE

Atrazine fate and transport within the coastal zone in southeastern Puerto Rico.

JOURNAL

Marine Pollution Bulletin

SIGNIFICANCE

Under certain conditions, atrazine concentration close to ecological effects thresholds was detected in the waters of Jobos Bay National Estuarine Research Reserve.

The atrazine was shown to enter these estuaries through surface runoff from adjacent agricultural land.

Management officials should consider the type of herbicide and the timing of its application to agricultural systems in coastal watersheds that drain directly into ecologically sensitive marine systems.

SUMMARY

This work examined atrazine fate and transport from a farm field adjacent to the Jobos Bay National Estuarine Research Reserve. Results showed that residues may move to estuary in shallow groundwater seepage and surface runoff. The predominant pathway appeared to be surface runoff. Atrazine concentration close to ecological effects thresholds was detected in one nearshore estuary water sample due to runoff during a tropical storm. After subsequent runoff events, atrazine and its two dealkylated degrades (DIA and DEA) were near or below detection limits in estuary samples. Contributing factors were runoff timing relative to atrazine application and rapid atrazine soil dissipation (DT50=1-3 days). Results indicated that adapted degradation conditions had developed. To improve weed management, atrazine replacement with other herbicide(s) is recommended. If active ingredient(s) selected have greater soil persistence runoff risks may be increased

ADDITIONAL ARTICLE

AUTHORS

Richard D. Methot Jr. (NOAA NMFS), Chantell R. Wetzel (UW)

PUBLICATION DATE

Early 2013 in print

TITLE

Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management

JOURNAL

Fisheries Research*

SIGNIFICANCE

Stock Synthesis supports assessment for a growing number of fish stocks, with 35 stocks in the US and 27 in non-US settings through 2011. Highly publicized controversies over assessment results highlight the importance of putting sufficient effort into building the best tools possible to take advantage of the great wealth of fishery and survey data being collected today.

Novel applications of Stock Synthesis are described for diverse species, such as tuna and dogfish shark, and may be used to investigate climate effects on fish productivity, and demonstrate its capability to work in data-poor situations and to estimate uncertainty in assessment results.

Dogfish Shark

SUMMARY

Stock synthesis (SS) is a statistical age-structured population modeling framework that has been applied in a wide variety of fish assessments globally. The framework is highly scalable from data-weak situations where it operates as an age-structured production model, to complex situations where it can flexibly incorporate multiple data sources and account for biological and environmental processes. SS implements compensatory population dynamics through use of a function relating mean recruitment to spawner reproductive output. This function enhances the ability of SS to operate in data-weak situations and enables it to estimate fishery management quantities such as fishing rates that would provide for maximum sustainable yield and to employ these rates in forecasts of potential yield and future stock status. Complex model configurations such as multiple areas and multiple growth morphs are possible, tag-recapture data can be used to aid estimation of movement rates among areas, and most parameters can change over time in response to environmental and ecosystem factors. SS is coded using Auto-Differentiation Model Builder, so inherits its powerful capability to efficiently estimate hundreds of parameters using either maximum likelihood or Bayesian inference. Output processing, principally through a package developed in R, enables rapid model diagnosis. Details of the underlying population dynamics and the statistical framework used within SS are provided.

ATTACHMENT

PDF: See attached.

*The December 2012 issue is devoted to articles about fishery stock assessment modeling using the Stock Synthesis software developed by NOAA researcher Dr. Richard Methot. Other articles by NOAA scientists are included in this issue.

ADDITIONAL ARTICLE

AUTHORS

William Rowland (NOAA/NESDIS/NGDC/STP), Robert Weigel (George Mason University, School of Physics, Astronomy, and Computational Sciences)

PUBLICATION DATE

November 21, 2012

TITLE

Intracalibration of particle detectors on a three-axis stabilized geostationary platform

JOURNAL

Space Weather

SUMMARY

We describe an algorithm for intracalibration of measurements from plasma or energetic particle detectors on a three-axis stabilized platform. Modeling and forecasting of Earth's radiation belt environment requires data from particle instruments, and these data depend on measurements which have an inherent calibration uncertainty. Pre-launch calibration is typically performed, but on-orbit changes in the instrument often necessitate adjustment of calibration parameters to mitigate the effect of these changes on the measurements. On-orbit calibration practices for particle detectors aboard spin-stabilized spacecraft are well established. Three-axis stabilized platforms, however, pose unique challenges even when comparisons are being performed between multiple telescopes measuring the same energy ranges aboard the same satellite. This algorithm identifies time intervals when different telescopes are measuring particles with the same pitch angles. These measurements are used to compute scale factors which can be multiplied by the pre-launch geometric factor to correct any changes. The approach is first tested using measurements from GOES-13 MAGED particle detectors over a 5-month time period in 2010. We find statistically significant variations which are generally on the order of 5% or less. These results do not appear to be dependent on Poisson statistics nor upon whether a dead time correction was performed. When applied to data from a 5-month interval in 2011, one telescope shows a 10% shift from the 2010 scale factors. This technique has potential for operational use to help maintain relative calibration between multiple telescopes aboard a single satellite. It should also be extensible to inter-calibration between multiple satellites.

ADDITIONAL ARTICLE

AUTHORS

Rong Zhang, Thomas L. Delworth, Rowan Sutton, Dan Hodson, Keith Dixon, Isaac Held, Yochanan Kushnir, John Marshall, Yi Ming, Rym Msadek, Jon Robson, Tony Rosati, MingFang Ting, Gabriel Vecchi

PUBLICATION DATE

January 2013.

TITLE

Have Aerosols Caused the Observed Atlantic Multidecadal Variability?

JOURNAL

Journal Atmospheric Sciences

SIGNIFICANCE

The authors show that the aerosol effects simulated in HadGEM2-ES cannot account for the observed anti-correlation between detrended multidecadal surface and subsurface temperature variations in the tropical North Atlantic.

These discrepancies cast considerable doubt on the claim that aerosol forcing drives the bulk of this multidecadal variability.

SUMMARY

Identifying the main drivers of the twentieth-century multidecadal variability in the Atlantic Ocean is crucial for predicting how the Atlantic will evolve in the coming decades and the resulting broad impacts on weather and precipitation patterns around the globe. Another recently published paper suggested that aerosols are a prime driver of twentieth-century North Atlantic climate variability, based on simulations using the HadGEM2-ES (UK Met Office Hadley Centre Earth System Model). In this paper, GFDL authors and colleagues show that there are major discrepancies between the HadGEM2-ES simulations and observations in the North Atlantic upper ocean heat content, in the spatial pattern of multidecadal SST changes within and outside the North Atlantic, and in the subpolar North Atlantic sea surface salinity. These discrepancies may be strongly influenced by, and indeed in large part caused by, aerosol effects.

AUTHORS

Rod Ehler

PUBLICATION DATE

~February 15, 2013

TITLE

Commercial Fishing in the Stellwagen Bank National Marine Sanctuary: Total Economic Value Baseline Assessment

SIGNIFICANCE

This analysis establishes the baseline socioeconomic value and importance of commercial fishing in Stellwagen Bank National Marine Sanctuary that can be used in future policy proposals for the region.

SUMMARY

This document is a social and economic assessment of commercial fisheries landings originating from Stellwagen Bank National Marine Sanctuary (SBNMS). The first step is the analysis of total socioeconomic value is to establish the baseline relationships between the local/regional economies and the use of SBNMS. The primary data source for the analysis is mandatory Fishing Vessel Trip Reports (VTR). VTR data are routinely used in analyses of spatial distribution of fishing effort and catches in fishery management plans. The data were analyzed to document and typify the spatial distribution, landings value (ex-vessel, dockside sales paid to fishermen) and volume, and species composition representative of commercial fisheries in the sanctuary. Ex-vessel or landings value is the price paid to the fishermen upon direct sale of the fish landed. In the final step in the analysis we take the estimates of direct impacts and extend the analysis to the total economic and social impacts. For the economic impacts, we address the market economic impacts (e.g. ex vessel revenues from commercial fisheries and associated impacts on sales/output, income and employment of the local/regional economies of the study area. This final analysis utilizes NMFS Northeast Fisheries Science Center's Northeast Region Commercial Fishing Input-Output Model (NERIOM) to estimate total economic impacts.

AUTHORS

Dr. Vernon R. Leeworthy

PUBLICATION DATE

Expected Publication Date: March 2013

TITLE

Technical Appendix: Knowledge, Attitudes and Perceptions of Management Strategies and Regulations of the Gray's Reef National Marine Sanctuary by Users and Non-users of the Sanctuary, Version 2, Marine Sanctuaries Conservation Series ONMS-13-xx.

SIGNIFICANCE

The information provided will assist management of GRNMS and education and outreach staff in better understanding who their users are, what they care about, and how to improve communication with both users and non-users of Gray's Reef National Marine Sanctuary.

SUMMARY

This research is part of the Socioeconomic Research & Monitoring Program for NOAA's Office of National Marine Sanctuaries. In 2010, a baseline study on the knowledge, attitudes and perceptions of Gray's Reef National Marine Sanctuary's (GRNMS) management strategies and regulations of users and non-users of GRNMS was initiated. Mail surveys of both users and non-users were designed in 2010 and implemented in 2011 for versions 1 for users and versions 1 and 2 for non-users. Version 2 for users was implemented in 2012. The study provides baselines on users' and non-users' knowledge, attitudes and perceptions of GRNMS management strategies and regulations, but also provides information on Socioeconomic/demographic profiles, activity participation and use in coastal & ocean waters off the Georgia coast both inside and outside GRNMS, sources of information on GRNMS used and the trust of sources used, familiarity with GRNMS rules and regulations, and attitudes about selected management strategies for coastal & ocean resources both inside and outside GRNMS.

continued

SUMMARY

Support for Selected Management Strategies for Coastal & Ocean Resources off the Coast of Georgia Inside versus Outside GRNMS:

1) Use of Marine Zoning: Non-users were much more supportive of the use of marine zoning in the ocean and coastal areas off the coast of Georgia than users. Non-users overwhelmingly supported the approach with about 76% responding yes, while users were overwhelmingly against with more than 60% responding no.

2) Use of Marine Reserves (no-take areas): Non-users were much more supportive of the use of marine reserves both outside and inside GRNMS than users. About 82% of non-users either strongly or somewhat supported marine reserves outside GRNMS and about 81% either strongly or somewhat supported marine reserves inside GRNMS. In stark contrast, 78.57% of users either had no support at all or were somewhat against marine reserves both outside and inside of GRNMS.

3) Research Only Areas: As with all zoning strategies, non-users were much more supportive of the use of research only areas both outside and inside GRNMS than users. More than 80% of non-users either strongly or somewhat supported both the use of research only areas outside and inside GRNMS. Again in stark contrast, more than 78% of users either had no support at all or were somewhat against the use of research only areas both outside and inside GRNMS.

4) Multi-species Fishery Management: Non-users were more supportive of this approach to fishery management than users. But neither group had a majority supporting this approach.

A majority of non-users were neutral (52.86%) and a plurality (38.63%) of users either had no support at all or were somewhat against this approach. More than 39% of non-users either strongly or somewhat supported this approach, while 34% of users either strongly or somewhat supported this approach.

5) Ecosystem-based Approach to Management of Coastal and Ocean Resources: Again, non-users were more supportive of this approach than users. About 60% of non-users either strongly or somewhat supported this approach, while only about 20% of users either strongly or somewhat supported this approach. A majority of users (54.55%) either had no support at all or were somewhat against this approach.

AUTHORS

Teal, J.M., R. Best, J. Caffrey, C.S. Hopkinson, K.L. McKee, J.T. Morris, S. Newman and B. Orem. Edited by **A.J. Lewitus, M. Croom, T. Davidson, D.M. Kidwell, B.A. Kleiss, J.W. Pahl,** and **C.M. Swarzenski.**

PUBLICATION DATE

To be posted on the web in early to mid-January 2013

TITLE

Mississippi River Freshwater Diversions in Southern Louisiana: Effects on Wetlands Vegetation, Soils, and Elevation

SIGNIFICANCE

The draft Position Paper has raised the state of Louisiana's awareness that Freshwater Diversions are not an effective approach for maintaining or restoring wetlands, and as a result, plans for the few Freshwater Diversions included in the Master Plan are being reconsidered. This is particularly important at this time when significant decisions are about to be made regarding diversion projects in Louisiana from BP settlement funding.

Statistical analyses to support guidelines for marine avian sampling: final report.

SUMMARY

Although Freshwater Diversions typically move some amount of sediments, the Panel was asked to evaluate the responses of marsh vegetation and soils to the effects of low salinity, high nutrient water in isolation. This information is not only important to assessing the effectiveness of existing and planned Freshwater Diversions that have intended consequences of promoting marsh sustenance and restoration by nourishing vegetative growth, but to informing model predictions of the influence of river water flux in all diversion types.

Agreement could not be reached on whether Freshwater Diversions would be beneficial or detrimental to marsh restoration over long term – too few representative systems have been studied, ecologically relevant designs are lacking, and measurements of some key indices of vegetation and soil responses are rare. Existing Freshwater Diversions have not been effective at restoring marshes – any benefits are compromised by hurricanes and other extreme events (prolonged flooding); Major amounts of sediment need to be moved for significant marsh restoration to occur. Science is needed on assessing how much sediment is effective for each case.

Good news is that most of the planned diversions in the LA Master Plan are Sediment Diversions, but question remains – is enough sediment being moved in each case?

Adaptive Management is key and LA CPRA has intentions for Adaptive Management as an important framework for moving forward with LA Master Plan – LA's feet need to be held to fire to ensure this is a priority.

AUTHORS

Brian P. Kinlan, Elise F. Zipkin, Allan F. O'Connell, and **Christopher Caldow**

PUBLICATION DATE

Expected publication date: December 31, 2012

TITLE

NOAA NOS NCCOS Technical Memorandum 158 and DOI OCS Study BOEM 2012-101

SIGNIFICANCE

We have developed and illustrated a simple, general method for defining species-specific hotspots and coldspots of occurrence and abundance in marine birds, and for assessing the significance and statistical power to detect these hot and coldspots. Given information about a species' regional occurrence and abundance patterns, this method can serve as the basis for general guidelines for the design of robust surveys to detect departures from regional average patterns of abundance and occurrence.

Taken together, the results of this study represent a methodology for a) using existing marine bird survey data to assess the state of knowledge about relative hotspots and coldspots of marine bird abundance and occurrence in offshore areas, b) planning future marine bird surveys in offshore areas to leverage existing data, and maximize probability of detecting any hotspots/coldspots of abundance/occurrence probability that may exist in discrete spatial planning blocks, and c) distributing sampling effort in time to ensure adequate representation of environmental and ecological variance.

SUMMARY

A practical approach is described to estimate the amount of sampling effort required to have sufficient statistical power to identify species-specific "hotspots" and "coldspots" of marine bird abundance and occurrence in an offshore environment divided into discrete spatial units (e.g., lease blocks), where "hotspots" and "coldspots" are defined relative to a reference (e.g., regional) mean abundance and/or occurrence probability for each species of interest. For example, a location with average abundance or occurrence that is three times larger the mean (3x effect size) could be defined as a "hotspot," and a location that is three times smaller than the mean (1/3x effect size) as a "coldspot." The choice of the effect size used to define hot and coldspots will generally depend on a combination of ecological and regulatory considerations. A method is also developed for testing the statistical significance of possible hotspots and coldspots. Both methods are illustrated with historical seabird survey data from the USGS Avian Compendium Database.

G Model
FISH-3547; No. of Pages 14
ARTICLE IN PRESS
Fisheries Research xxx (2012) xxx–xxx
Contents lists available at SciVerse ScienceDirect
Fisheries Research
journal homepage: www.elsevier.com/locate/fishres

Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management

Richard D. Methot Jr.[a,*], Chantell R. Wetzel[a,b]

[a] Office of Science and Technology, National Marine Fisheries Service – NWFSC, 2725 Montlake Boulevard East, Seattle, WA 98112, United States
[b] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, United States

ARTICLE INFO

Article history:
Received 11 May 2012
Received in revised form 11 October 2012
Accepted 30 October 2012

Keywords:
Stock assessment
Stock synthesis
Population models
Fishery management

ABSTRACT

Stock synthesis (SS) is a statistical age-structured population modeling framework that has been applied in a wide variety of fish assessments globally. The framework is highly scalable from data-weak situations where it operates as an age-structured production model, to complex situations where it can flexibly incorporate multiple data sources and account for biological and environmental processes. SS implements compensatory population dynamics through use of a function relating mean recruitment to spawner reproductive output. This function enhances the ability of SS to operate in data-weak situations and enables it to estimate fishery management quantities such as fishing rates that would provide for maximum sustainable yield and to employ these rates in forecasts of potential yield and future stock status. Complex model configurations such as multiple areas and multiple growth morphs are possible, tag-recapture data can be used to aid estimation of movement rates among areas, and most parameters can change over time in response to environmental and ecosystem factors. SS is coded using Auto-Differentiation Model Builder, so inherits its powerful capability to efficiently estimate hundreds of parameters using either maximum likelihood or Bayesian inference. Output processing, principally through a package developed in R, enables rapid model diagnosis. Details of the underlying population dynamics and the statistical framework used within SS are provided.

Published by Elsevier B.V.

1. Introduction

Fishery stock assessment models are demographic analyses designed to determine the effects of fishing on fish populations and to evaluate the potential consequences of alternative harvest policies. The conceptual framework for these analyses is provided by simple models of the demographic processes birth, natural and fishing death, growth, maturation, and movement. However, it is rare that sufficient information on all these processes is available over the time period during which a fishery has affected the stock in question. The history of fish stock assessment models was dominated in early years by two, quite different approaches to this problem. One branch of investigation used a time series of an indicator of stock abundance, typically based on the standardized catch rate in the fishery as a proxy for stock abundance, and the time series of fish catch to calibrate a simple two-parameter production model (e.g. Schaefer, 1954) that could provide inference about current and target fish stock abundance and the maximum sustainable yield. The other approach, cohort analysis or virtual population analysis (VPA), depended on a time series of detailed fishery catch-at-age data to reconstruct the virtual abundance of each annual cohort that had been fished (Pope, 1972; Laurec and Shepherd, 1983). This virtual population analysis approach works best when fishing mortality rates have been higher than natural mortality rates, but only reconstructs the historical abundance and fishing mortality rates. Population productivity is then estimated by fitting a functional relationship between the expected value of annual recruitment and spawning stock abundance.

Over the past 20 years, there has been development of a third approach to the fishery analysis problem. Integrated analysis (IA) (Maunder and Punt, in this volume) takes a more inclusive approach to modeling the population dynamics and utilizing a wide range of available data. A definitive beginning for IA is the landmark paper by Fournier and Archibald (1982). The stock synthesis (SS) implementation of IA began during the early 1980s (Methot, 1986, 1989) following Fournier's pioneering work and adopted many of the characteristics of his approach.

Why "stock synthesis"? The term synthesis is used in the context of development of a new product that is more than an amalgam of its disparate parts. In fish stock assessments, different kinds of data can provide complementary information about the stock of fish, but one source may not be sufficient in itself to provide a complete picture of the stock's abundance and the impact of fishing on the stock. SS inherently blends the population estimation paradigm

* Corresponding author. Tel.: +1 206 860 3365; fax: +1 206 860 3394.
E-mail address: richard.methot@noaa.gov (R.D. Methot Jr.).

0165-7836/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.fishres.2012.10.012

Please cite this article in press as: Methot Jr., R.D., Wetzel, C.R., Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. (2012), http://dx.doi.org/10.1016/j.fishres.2012.10.012

of virtual population analysis with the population productivity paradigm of biomass dynamics models. The population dynamics sub-model of SS is quite comparable to the population dynamics found in most statistical catch-at-age models. However, the observation sub-model of SS is highly developed to use a wide range of available types of data to calibrate the model. The observations that can be included in SS include: fishery CPUE or effort; survey abundance; discards; length-, age- and weight-composition data; and tag-recapture data. Technical details are provided in Appendix A to update Methot (2000). The estimation of target fishing mortality rates and application of these rates in forecasts of potential fishery yield and stock status have been developed extensively in SS. SS evolved as much from an ecological perspective as from a statistical perspective. Thus, it includes capability to use time series of environmental and ecosystem factors to influence the population dynamics and observation processes over time.

SS has evolved and grown to meet the needs of the fishery stock assessment community over the past near 30 years. 35 stocks in the US, 10 tuna/billfish in three oceans, four European stocks, and 12 Australian stocks have been assessed using this approach by 2012 (Appendix B). Major milestones in this history are recapped in this paper and the current set of SS features is described. Updates to the technical description will be made available on the web to document the evolving features of SS.

2. History

The stock synthesis assessment approach has progressed through three major stages of development since its beginning 30 years ago. These stages include an initial development for northern anchovy, *Engraulis mordax*, in which the basic concept was established, re-development as a generalized model focused on west coast groundfish, and a third stage in which the computer code was translated to ADMB (Fournier et al., 2011) to take advantage of the power of automatic differentiation.

2.1. Anchovy synthesis

The original model for northern anchovy (Methot, 1986, 1989) melded several diverse data sources. Age-composition data were available from the fishery, but fishing mortality rates were low and aging imprecision was a factor, so virtual population analysis could not be used. An egg and larval survey provided a time series of relative spawning biomass, but surveys were triennial during much of the time period, thus hampering a production model approach. A new Egg Production Method had been developed (Lo et al., 1992) to provide a fully calibrated, direct measure of spawning biomass, but only a few estimates by this method had been conducted by the early 1980s. The anchovy synthesis model integrated these diverse data types into an age-structured history of the anchovy stock off California.

The anchovy SS model incorporated several factors related to the biology of this fish. A latitudinal gradient in age composition and fluctuating ocean conditions that shifted this gradient north-south exposed a varying component of the stock to the relatively stationary fishery. Time-varying fishery selectivity was incorporated in the model to deal with this phenomenon. Second, biological studies demonstrated a temperature dependence on anchovy maturity (Parrish et al., 1986). Thus, inclusion of environmental effects (ocean temperature) on biological factors (age-specific maturity) in the model helped to explain the dip and then rapid resurgence of the larval index during the early 1970s. Third, anchovy are a forage fish and a known predator, Pacific mackerel *Scomber japonicus*, was changing rapidly in abundance. Thus, time-varying natural mortality as a function of an external ecosystem driver, the abundance of Pacific mackerel, was a feature of the model. These concepts of time-varying factors linked to environmental and/or ecosystem factors have remained in SS ever since. Unfortunately, situations in which there has been sufficient information to invoke these features have been rare.

A bit of history is worth noting at this point. The anchovy SS was built at the dawn of the personal computer era. It was written in FORTRAN for hand-built computers termed the "LaJolla Standard" running the CP/M operating system with 8-in. floppy disks as the storage medium. Overnight model runs were normal and the code was laced with line printer statements so that something could be retrieved after various failures. In hindsight, as computers have become orders of magnitude faster, the complexity of our modeling approaches have kept pace so that weekend long Markov Chain Monte Carlo (MCMC) runs now represent the frontier of our patience for waiting for model results.

2.2. Moving to groundfish

The second stage of SS evolution began in 1988 as NOAA's Alaska Fisheries Science Center sought improvement to the modeling framework for west coast groundfish, particularly sablefish (*Anoplopoma fimbria*) and Pacific hake (whiting) (*Merluccius productus*).

The principle challenge for sablefish was to find a way to take advantage of the new, but growing, time series of fishery and survey size-composition data. Aging data were scarce and fraught with aging imprecision and bias. The modeling concept developed to deal with this situation treated all composition data as a proxy for the actual, true age-composition of the fishery catch or a survey. A matrix in the model transformed the estimate of true age-composition into an expected value for the composition as measured. The transformation matrix could be a simple 1:1 relationship if the fish aging was asserted to be perfect, could be an age matrix that included aging imprecision and/or bias, or could be a matrix that transformed the age-composition into a length-composition estimate using the growth curve (Methot, 1990). Also, sablefish were caught with fishing gear that was size-selective for larger sablefish, but older sablefish were known to diffuse into deeper water out of the range of most fisheries, so the sablefish SS model incorporated both size- and age-selectivity that could be applied simultaneously. Finally, use of the size-composition data required a growth curve, but external estimates of the growth curve would have been biased by the inability to account for size-selectivity and aging imprecision. Thus, the sablefish SS model included the ability to estimate growth parameters while accounting for the influence of size-selectivity and aging imprecision. Data that enabled estimation of growth parameters included modes in the size-composition data and observed mean size-at-age data.

The situation for Pacific hake was quite different than for sablefish. Here there were complete age-composition data from the fisheries and surveys extending nearly 20 years and a history of modeling using VPA. The interest was in exploration of alternative modeling approaches, especially approaches that could address the strong inter-annual shifts in stock abundance between the US and Canada in response to el nino – la nina ocean conditions. An SS model was constructed using only age-selectivity in the fishery process. Its innovation was in allowing for multiple stock areas with annually varying proportion of the stock in each area (Methot and Dorn, 1995).

Since the inception of the sablefish and hake SS models in 1988 and 1989, their use expanded beyond sablefish and hake to most Pacific coast groundfish by the late 1990s. As this usage grew, ADMB was being used by numerous researchers to create other statistical catch-at-age models. The SS models were coded in FORTRAN and used numerical derivatives to obtain the gradient information

Please cite this article in press as: Methot Jr., R.D., Wetzel, C.R., Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. (2012), http://dx.doi.org/10.1016/j.fishres.2012.10.012

needed to iteratively search for the best set of model parameters. The analytically calculated derivatives available through ADMB offered faster and better model performance.

2.3. Reborn in ADMB

The SS transition to ADMB provided an opportunity to merge the features of the FORTRAN-coded models for sablefish and hake into a fully generalized framework. SS then inherited the powerful features of ADMB for rapid model convergence and variance estimation through inverse Hessian or Bayesian MCMC approaches. This model, termed SS2, began to be used for west coast groundfish assessments in 2005. Evolution of SS2 was rapid and a major re-work of some internal structures in 2009 led to the awkward renaming as SS3.

A note on SS nomenclature and distribution is worth including here. Today, the model executable code is labeled SS3, but the model framework itself is generically referred to as SS. Upgrade versions are tracked with a specific nomenclature. In 2012, SS is at V3.24f where the 3.2 refers to a major model update involving mandatory changes to input files, the ".x4" refers to the addition of some specific feature usable in particular circumstances, and the "f" keeps track of minor tweaks and fixes. The history of model updates is distributed with SS as an annotated, filterable list, and the specific version information and compile date are output at the top of all SS output files.

3. Overview of features

The detailed features and formulas found in SS are described in Appendix A. The sections below provide a brief overview of major features of SS: stock-structure, spawner–recruitment, life history and biology, selectivity, fishing mortality, observation sub-model, fishery management targets and forecasting, variance estimation, and output processing.

3.1. Stock structure

The total population can be divided among one to many biological entities, and the numbers-at-age of each entity are tracked over time. Some of these can have unique biology (gender, growth and natural mortality) and some can have a unique season of birth within a year. The total of all entities born within a year is referred to as a year-class or cohort. Each of the biologically or birth season-delineated entities will be referred to as a morph. Each morph can be divided into males and females with gender-specific growth and natural mortality rates. In addition, each morph can be sub-divided into slow-, medium- and fast-growing entities termed platoons (Goodyear, 1997). The phenomenon of platoons within a morph is not observable. However, the logic of their existence is as biologically simple as the concept that a fish that is near the upper end of the size-at-age distribution one year will be within that upper distribution during the following year. Expected values for all observations are based on summation over all platoons and morphs, except male and female data can be kept disaggregated.

To the extent that size-selective fisheries are implemented, each slow-, medium- and fast-growing platoon will have unique age-specific mortality and survivorship (Taylor and Methot, in this volume). The age-specific mortality on a platoon is calculated according to the average selectivity-at-age, which is the dot product of selectivity-at-size and the normal distribution of size-at-age for that platoon. Thus, there can be differential survival between morphs with different growth rates and between platoons within a morph.

In a multiple area model configuration, each entity can be distributed across multiple areas at birth and move between areas according to the unique age-specific movement rates of their morph. The model parameters that control the proportion of the total number of births to each morph and subsequent movement rates are potentially estimable with sufficient data and, like all parameters in SS, can be defined to change over time. If the morph feature is being used with platoons, then the platoons move identically to their parent morph. Size- and age-specific mortality is applied within each area to each entity occurring in that area.

In practice, most SS configurations have just two morphs, one for females and one for males, and only one platoon per morph. More complex situations can be created to more closely match biological reality. For example, a model with two morphs, three areas and four seasons could have one morph recruit in area 1 and the other morph recruit in area 2, then both morphs migrate seasonally into area 3 and then back to their natal area in a later season. Models with five areas have been created for bigeye tuna *Thunnus obesus* (Aires-da-Silva and Maunder, 2012). Another morph-related concept is hermaphroditism. Just as fish of a given platoon can move between areas at the end of each time step, the hermaphroditic function allows an estimable fraction of the females to transfer into the corresponding male platoon at the end of the time step.

3.2. Spawner–recruitment

The spawner–recruitment function defines the linkage between the reproductive potential (total fecundity) of females and the expected total number of age zero animals. In a single gender model, all fish are implicitly female; there is no division into males and females and only total abundance is modeled and reported. Male mature biomass is included in the calculation of reproductive potential in a two-gender model only when hermaphrodicity is accounted for. The Beverton–Holt spawner–recruitment function is most commonly used and other available functions include Ricker, hockey stick and a three-parameter survivorship-based function (Taylor et al., in this volume). In a multi-area model, the stock's reproductive potential is summed over all areas to create a global spawner–recruitment relationship. All female morphs and platoons contribute to the global reproductive potential which will create the next global recruitment of a cohort, which is then distributed among the platoons and morphs. There is no heritability to carry the surviving proportions among platoons and morphs into the next generation.

The output of the spawner–recruitment function is the expected mean number of age-0 animals, not the median number. When lognormal recruitment deviations are applied, they are added to a downward bias adjustment from the logarithm of this expected mean recruitment to account for the expected difference due to the variability among the estimated recruitments (Appendix A, Eq. (A.1.7)) (Methot and Taylor, 2011). The magnitude of this bias adjustment offset can change during the time series being analyzed. In maximum likelihood estimation, it should approach zero in data-poor historical eras with insufficient data to estimate the time sequence of recruitment deviations. The offset should approach the maximum only where there are sufficient data to estimate the time sequence of recruitment deviations with high precision. Because the population's biomass is dependent on the mean recruitment, not the median recruitment, this offset approach assures that the expected biomass will not change as a consequence of having more data from which to estimate the true recruitment deviations (Methot and Taylor, 2011; Stewart et al., in this volume). In MCMC estimation, the bias adjustment is set to 1.0 for all years because the estimation procedure will integrate over the full lognormal range of potential recruitments.

The duration of the life stage between spawning and recruitment to the population is not explicitly modeled, so the age 0 animals instantly exist as early as the start of the time step in which

spawning occurs. The cohort begins to experience age-specific natural and fishing mortality and to grow according to the defined growth function at this time. In a multi-season model, the total number of age 0 animals can be distributed among seasons starting with the season in which spawning occurs and ending with the last season before the next spawning event. Thus, some of the age 0 animals have their age 0 event delayed until the start of a subsequent season. These later-born morphs are identified by their birth season. Even if these later-born morphs have the same growth parameters as early-born morphs, their delayed age zero event will cause them to be smaller at calendar age than the earlier-born morphs. However, for the purposes of age determination, all fish born in a calendar year are considered to have the same integer age and to graduate to the next integer age on January 1 of the following year. Care must be taken when assembling data for input to SS to assure that the manner in which integer ages are assigned by otolith readers matches this convention in the modeling of age progression.

3.3. Life history/biology

Fish body weight is needed to convert the modeled numbers-at-age into quantities such as total catch biomass or a biomass-based abundance index. SS provides an approach for inputting empirical body weight-at-age observations, but typical applications generate estimated body weight-at-age from biological processes in the model. The empirical body weight-at-age approach is common in statistical catch-at-age models and appropriate when there is high precision in these data. However, if these data are not of high precision, the variance in these empirical body weights-at-age should contribute to variance in model estimates. SS provides the capability to calculate body weight-at-age from a length-at-age growth function for each morph, which itself is based on estimable parameters, a body weight-at-length function, and size-selectivity of the fishery to which that body weight-at-age pertains. This capability was critical for the first SS models for Pacific coast sablefish where age data were uncertain. In this approach, SS calculates a different body weight-at-age for the retained fish versus the discarded fish of each fishery for which a size-based retention function is applied. SS calculates mean body weight at true age for use in population dynamics. Mean body weight (or length)-at-age data can be input to SS for model calibration and will inform the estimation of growth parameters while taking into account effect of size selectivity and aging imprecision. Differences in body weight data between fleets will influence estimation of size selectivity for these fleets. Differences in body weight data between retained and discarded fish will influence estimation of the size-retention function. The full integration of growth estimation, size selectivity and retention estimation, and aging imprecision is a major strength of SS.

Body growth in length is calculated according to a growth function, typically the von Bertalanffy. When the model configuration incorporates time-varying growth, the current year's parameters are applied to the current size-at-age of each biological entity to calculate its growth increment toward the current asymptotic size for that morph. Thus, growth is according to morph and fish cannot shrink even if the asymptotic size becomes reduced below their current size, although in the hermaphroditic model configuration the gender change causes an immediate switch from female size-at-age to male size-at-age so care must be taken in the setup of the growth curves. Another model feature allows for cohort-specific growth rates (but not cohort-specific asymptotic size) which could be relevant if there is density-dependent growth according to relative cohort abundance. Fish of each morph can be distributed across areas in a multi-area model. The growth characteristics are associated with the morph, not with the area in which they reside. However, it is reasonable to create a configuration where each area has its own predominant morph that then has some degree of movement to other areas. Expected values for data, such as the overall size composition of a fleet's catch in an area, take into account the combined catch across all morphs.

SS follows Schnute (1981) re-parameterization of the von Bertalanffy growth curve to redefine the growth parameters in terms of size at two reference ages. Below the lower reference age, where there presumably is little information about the actual size-at-age trajectory, SS simply applies a linear trend to grow the fish from a small size at the age zero event until they reach the lower reference age of the growth curve. More flexibility in the trajectory of growth can be achieved by using the three parameter Richards growth function, or even more by invoking age-specific growth rates for each of several younger ages.

Age-specific natural mortality, maturity, and fecundity can follow various functional forms or be input empirically (Appendix A). All of these are defined as model parameters, but only natural mortality can be treated as estimable because there is currently no capability to include maturity-at-size or fecundity data in a manner that would allow internal estimation of functions related to these processes.

3.4. Selectivity

SS provides a range of options for calculating the age-, size- and gender-selectivity of each fishing fleet and survey. SS distinguishes fishing fleets from surveys by specifying that catch is input for fishing fleets, thus the fleet's catch affects the population, whereas a survey is for observation only and does not affect the population dynamics. Most options define a fishery's or survey's selectivity by a smooth parametric function, rather than an age-by-age approach. In common with growth, this approach was implemented in the early model for sablefish where size-composition data were sparse and age-composition data nearly non-existent. The sablefish situation also required the capability for a fleet's selectivity to have age-, size-, and gender-dependency. Size selectivity against small sablefish was logical as they recruited to the longline and pot gear, and declining age selectivity was logical as older sablefish diffused into deep water away from the fishing grounds.

SS includes over a dozen size- and age-based selectivity functions as defined in Appendix A, Section 5.6. These range from simple logistic, to double logistic and double normal, and typically have 2–8 free parameters. The simpler the parametric form for the selectivity, the more the assessment result is constrained by that parametric structure. Similarly, when assessment configurations assert that fishery selectivity is unchanging over time, this is a strong assertion that the fishing fleet is a good sample of the population's size/age composition as filtered by that time-invariant selectivity. Sampson and Scott (2012) provide a sound rationale for doubting time-invariant fishery selectivity. Taylor and Methot (in this volume) show how time-varying selectivity can be invoked in SS to more fully explore a plausible range of model outcomes. For example, selectivity could be near asymptotic during early years when older/larger fish were more abundant, then transition to a more dome-shaped pattern as fishing reduces the abundance of older/larger fish and thus reduces the fraction of the fishing effort deployed in habitats where older/larger fish typically are found. A variety of options are available to allow selectivity (and other) model parameters to change over time. This can be according to user-specified blocks of years, random deviations in selectivity over a range of years, random walk changes over a range of years, a time trend based on a cumulative normal distribution, and as a function of an auxiliary time series, termed an "environmental index". These options are portrayed visually in Fig. 1.

Non-parametric age-selectivity can be used when the maximum age is relatively low and there are reasonably good age data.

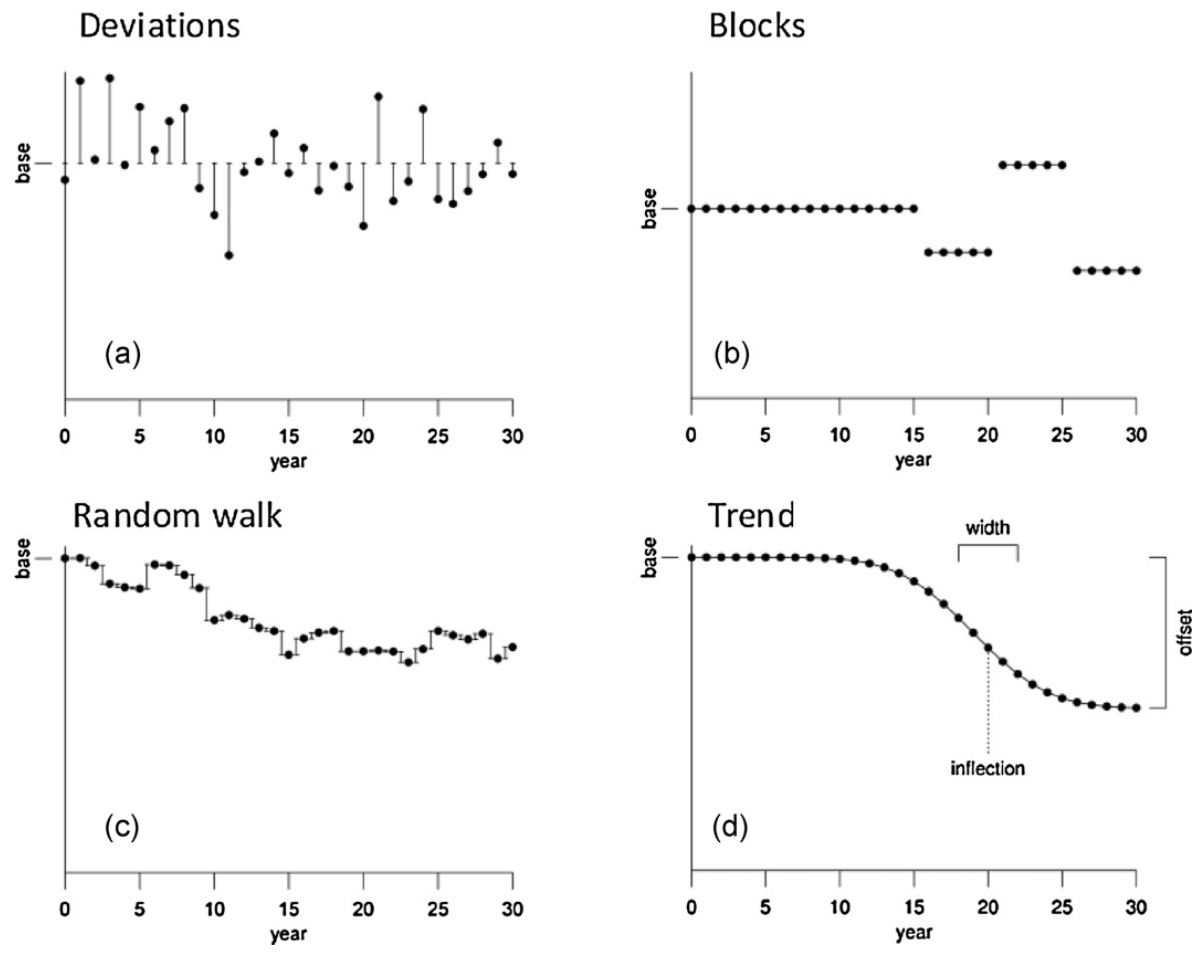

Fig. 1. Stock synthesis includes five options for allowing parameters to have time-varying values. Shown here are options for annual deviations (a), time blocks (b), annual random walk deviations (c), and Gaussian trend (d). Not shown is the fifth option which links the values for the parameters of the selection function to an environmental time series.

However, the flexibility of size selectivity is not so easily mapped by a natural granularity (large number of length classes implies many additional parameters) to the data. Even more serious an issue would be the very large number of parameters needed for time-varying non-parametric selectivity. The approach of Butterworth et al. (2003) to apply a selectivity surface smoothed in the age and time dimension works well when there are few fleets and reasonably complete data. However, parametric forms are necessary when there are many fleets and sparse data.

A cubic spine selectivity function (Fournier et al., 1998) achieves a reasonable compromise between non-parametric and parametric selectivity. SS can create starting values for the parameters of spline selectivity functions de novo from the characteristics of the data by aggregating each fleet's or survey's size- (or age-) composition data over all years, distributing the user-specified number of cubic spline knots across that cumulative probability distribution and creating parameter starting values. This capability emphasizes that SS is both a population dynamics model and a software system to implement that model.

3.5. Fishing mortality

SS was designed around an approach that relies on the absolute level of catch being known well enough to allow the model to calculate the level of fishing intensity needed to obtain that level of catch conditioned on the model's current estimate of age-specific population abundance and age-specific selectivity. Because the model first calculates catch-at-age in numbers for each fleet, it is straightforward to calculate the total catch for each fleet in terms of both numbers and weight and to provide the capability to use either as the metric in which a fleet's catch is entered into SS. Importantly, this also allows the forecast catch targets for each fleet to be in either numbers or weight. This is valuable for recreational fishery data which typically is available in numbers caught.

SS does not rely upon fishing mortality rate being estimated as catchability multiplied by fishing effort. Instead, seasonal, fleet-specific fishing intensity is directly estimated to match the observed catch. However, fishing effort and catchability can be brought into the calculations when data on fishery effort or catch per unit effort, CPUE, are provided for parameter estimation purposes. Here, the expected value for CPUE is estimated catchability multiplied by available biomass, and the expected value for fishing effort is 1/catchability multiplied by fishing intensity.

SS provides three approaches for the calculation of apical F for each fleet in each time step and area. These are Pope's approximation, which calculates the harvest rate needed to remove the catch at the midpoint of a season, continuous instantaneous F, with each F as a model parameter, and a hybrid approach (Appendix A, Section 1.10) which calculates continuous instantaneous F values through a differentiable iterative approach. The Pope and hybrid approach treat the harvest rates or F values as coefficients to be tuned within each model iteration to maintain the match between observed and expected catch, whereas the parameter-based F approach requires ADMB to estimate these F values as model parameters (after getting starting values using the hybrid approach in early model phases then converting over to the parameter approach in final stages). The use of the parameter approach to define fishing intensity means that the model will be able to account for uncertainty in the catch data and even allow for missing catch data if there is alternative information such as an effort time series.

A brief comparison of the performance of these three fishing mortality approaches in a simple catch-at-age model with 30 years of catch is informative (Table 1). The overall negative log likelihood and ending biomass are very similar between methods and

Table 1
Comparison among three approaches for calculating the level of fishing intensity in each time step of the model.

	Pope's	Parameters	Hybrid
Number of iterations	458	846	469
−log likelihood	1327.64	1327.68	1327.68
End year biomass	6017.2	5797.9	5798.1
Standard error of end year biomass	1331.5	1294.0	1293.9

the small difference for Pope's method is due to selectivity acting slightly different when used as a mid-year harvest versus continuous harvest (Branch, 2009). The similarity of the standard error for estimated spawning biomass (as determined by the inverse Hessian) is very reassuring, especially given that the parameter approach uses 30 parameters mapped to 30 *F* values, whereas Pope's and the hybrid method do this using internal coefficients. Although the parameter approach was slower (more iterations) in this example, model convergence can be slow for Pope's and the hybrid method in high *F* situations and the '*F* as parameter approach' is able to creep up on the final solution more quickly.

The term "fishing intensity" is used because the annual instantaneous rate of fishing mortality, *F*, across all fleets is not easily describable by a simple scalar, as it would be in a simple model with only a single fishery and with knife-edged selectivity to this fishery. In SS, there can be many fleets operating simultaneously, some in different geographic zones, and each with selectivity that could vary with age, size and gender. Thus, *F* has multiple dimensions. Three options are provided in SS for portraying the overall intensity of fishing: exploitation fraction, equilibrium reproductive output per recruit, and annual numbers weighted *F* over a range of ages. Exploitation fraction is simply the total annual catch divided by the total abundance above a specified age, with biomass- and numbers-based options provided. The equilibrium-based quantity is the spawner potential ratio (SPR) (Goodyear, 1993) calculated as the equilibrium level of spawning biomass-per-recruit (SPR) that would occur with the current year's level of fishing intensity relative to the unfished level of spawning biomass-per-recruit. This is reported as 1-SPR to create a metric that increases as fishing intensity increases. All seasonal, geographic, and selectivity issues are integrated into the calculation by focusing on the effect of fishing intensity on spawning biomass. The total *F* approaches the problem differently. Here the total annual fishing mortality for each age group (summed over all regions and calculated on a January to January basis) is calculated by:

$$F_{a,t} = \ln(N_{a+1,t+1}) - \ln(N_{a,t}) - M_a$$

where $N_{a,t}$ is population numbers-at-age, and M_a is natural mortality rate at age.

SS provides the capability to calculate this *F* statistic for a single age, or over a range of ages. The 1-SPR and total *F* statistics provide complementary information on the intensity of fishing; one on the cumulative impact of fishing on the entire stock and the other on the fraction of fish removed for key ages.

3.6. Expected values for data

Integrated analysis models seek to implement the goodness-of-fit calculations using data that are as lightly processed as possible (Maunder and Punt, in this volume). Some other models, particularly VPA, require a complete catch-at-age and body weight-at-age time series. In some cases these data are created by slicing size-composition data into bins representing the size range believed to represent an age of fish (Kell and Ortiz, 2011). In other cases, missing catch-at-age data are imputed through some interpolation protocol. Unfortunately, the variance associated with these procedures is not calculated and propagated into the assessment results. The impetus to create catch-at-age data sets has been so pervasive for some stocks in European waters that the details of the historical gap-filling have been lost and it may not be possible to apply a model such as SS to less processed data for these stocks (ICES, 2012).

There is no expectation in SS that data are in the form of catches-at-age, nor that data need to be available for all years of the analysis. Instead, SS takes the data in a lightly processed form and calculates an expected value while taking the various processes that are believed to have influenced the creation of each datum into account. For example, when mean body length-at-age data are input to SS, calculation of the expected value for these data takes into account: the estimated population size-at-age, size-selectivity of the fleet from which the sample was taken, aging imprecision which tends to blur information between adjacent ages of fish, and the estimated recruitment time series.

A powerful feature of SS is its ability to calculate expected values for a wide diversity of data types. Similar capability is found in the CASAL assessment program (Bull et al., 2005). The basic approach is as follows. In each time step, SS tracks the numbers-at-age for each entity by area. SS also tracks the mean and distribution of size-at-age for each entity through time. The vector of numbers-at-age then scales the distribution of size-at-age to create a matrix of numbers-at-size and -at-age for each morph in each area in each time step. Age/size/gender selectivity is then used to create a matrix of selected numbers at size/age for each fishery or survey fleet with data from that area/time. This matrix forms the basis for calculation of expected values for any data type. For each morph, the male and female matrices enable calculation of expected values that are either for combined genders, single genders, or joint genders with preservation of the sex ratio information. For a survey index, the entire selected matrix is simply summed to create the expected value to which catchability is then applied. The size-retention vector is applied to the matrix before calculating the expected values for either the retained or discarded components of the catch if the data are specific to retained versus discarded catch. The aging error matrix is applied before accumulating the expected values into the bins of observed ages if the data are ages. Thus, quantities such as the mean size-at-age of discarded fish can be calculated after accounting for aging imprecision. The various forms in which length, age, weight and abundance data can be represented in SS is outlined in Table 2. A more detailed technical description is found in Appendix A.

An example from the assessment for pink shrimp (*Fartantepenaeus duorarum*) in the Gulf of Mexico[1] illustrates the approach of fitting data in their native units. Previously, Nichols (1986) had assessed shrimp using a monthly cohort analysis. However, catch-at-month age data are not collected, in fact shrimp cannot be aged using conventional methods. The only composition data were the seven weight categories in which shrimp are sold to the processors. Nichols used a simulation model of shrimp growth and mortality to generate expected proportions at size and then used these proportions to slice the monthly weight composition data into monthly catch numbers-at-age, which were then analyzed using an untuned VPA. SS uses essentially the same process, but in reverse. Within SS, the simulation of recruitment, growth, selectivity and mortality using the parameter estimates leads to an estimate of catch numbers at monthly age. The application of the distribution of size-at-age from the growth curve results in catch numbers-at-length for each month. The generalized size frequency feature of

[1] Hart. Richard. 2012. Unpublished report. Stock Assessment of Pink Shrimp (*Farfantepenaeus duorarum*) in the Gulf of Mexico.

Table 2
The data types that can be included in SS.

Type	Units	Description
Catch	Biomass or numbers	For each fishing fleet. Initial catch calculation is numbers-at-age for each fleet, then can be aggregated to biomass or numbers. If *F* is estimated as parameters, then lower precision can be assigned to each catch value. Catch can be size-partitioned into retained and discarded portions, with *F* tuned to match the retained catch amount
Fleet or survey index	Biomass, numbers, effort	Is aggregated across ages/sizes according to specified or estimated selectivity, so age-specific index is feasible; catchability can be estimated as scaling constant or as a parameter, which can be time-varying or density-dependent; If units are effort, then expected value is F/q
Discard	Fraction, biomass or numbers	Created by applying a discard/retention logistic function to the total catch
Mean weight	Body weight	A simple measure of the mean size of fish in the catch (e.g. total weight/total numbers)
Length composition	Proportion	From each/any fleet/survey; combined gender, each gender, or split to preserve sex ratio; from total catch or discard portion or retained portion. Small constant can be added and tails can be compressed
Age composition	Proportion	As with length composition. Also, can be interpreted after applying aging error to expected values. Can be for entire length range, or for specified subset of length range (age-at-length approach)
Generalized size composition	Proportion	As with length composition, but with generalized bins which can be in terms of body length or weight, and accumulated in terms of numbers or biomass in each bin
Mean size-at-age	Body length or body weight	The expected values for these data take into account growth, size-selectivity, aging imprecision
Tag-recapture	Number of tags released and recaptured	Releases are for a specified age in a specified area at a specified time. Recaptures are from a particular fleet (which can only fish in one area) at a particular time
Morph composition	Proportion	In a set-up with multiple growth types (morphs), the proportion of each morph observed (e.g. by otolith micro-constituents) by a fleet/survey can be input
Environmental index	Time series index	Can be input as a "survey" (above) if used as data about recruitment deviations, or can be used as an input driver of a functional parameter relationship

SS then accumulates the catch weight at length into the bins of the catch weight categories to calculate an expected value for the observed catch weight categories. Variance in estimated recruitment, growth and mortality is based upon the goodness of fit to the actual weight composition data. SS is not misled into acting as if monthly catch-at-age data actually exist. The observation submodel in SS is configured to create expected values for the data, rather than forcing the data to be manipulated to match the structure of the model. This preserves the variance characteristics of the data and allows for estimation of the variance of the processes that created the data.

3.7. Fishery management targets and forecasts

SS calculates fishing intensity levels that would satisfy several common fishery management conventions. It does this using a dynamic pool approach (Mace and Sissenwine, 1993) that first calculates the equilibrium level of SPR and yield per recruit (YPR) that would occur if fishing according to a trial level of fishing intensity, then uses this SPR and the unfished level, SPR_0, to calculate the absolute level of recruitment, spawning biomass, and yield that would occur if fishing intensity were maintained at this rate. SS then iteratively calculates the level of fishing intensity that would match each of three conditions: (1) SPR – a user-specified target level of SPR relative to SPR_0, (2) B_{target} – a user-specified target level of equilibrium spawning biomass relative to the unfished level, and (3) MSY – the level of fishing intensity that would maximize yield. The latter two conditions take the spawner–recruitment relationship into account. These calculations utilize all the multi-fleet, multi-area, multi-season, size- and age-selectivity, complexity in the estimation model, so produces results that are entirely consistent with the assessment result.

SS, in common with many integrated analysis stock assessment models, is basically a simulation of a stock's age-structured population dynamics. This enables SS to utilize a selected fishing mortality approach (e.g. harvest policy) to extend into a forecast of the future age-structured stock abundance and yield that would occur while fishing according to that harvest policy (Maunder et al., 2006). SS allows the user to specify that forecasting use either the SPR-based fishing rate, the B_{target} rate, the MSY rate, or a multiple of the recent average rate when conducting the forecast.

The sequencing within SS calls the fishing mortality target routine and then the forecast routine in two circumstances. First, the routines are not called until convergence occurs and ADMB enters into the variance estimation phase when SS is operating in ADMB's estimation phase. Second, the routines are called during the MCMC evaluation phase so that the equilibrium and forecast results become part of the output for each selected set of parameters. This means that all aspects of parameter uncertainty calculated using the inverse Hessian method in the maximum likelihood estimation, and using the Markov Chain Monte Carlo approach are propagated into the variance of derived quantities, such as the fishing mortality intensity that would produce MSY, and forecasts of stock abundance and future yield that would occur while fishing according to the chosen level of fishing intensity.

Fishery management in the United States has strongly embraced an approach that utilizes catch quotas (termed annual catch limits), and expects that the annual catch limit be set according to a protocol that acknowledges scientific and management uncertainty. A technical approach has been developed to forecast future catch levels that would have a specified probability of allowing catch to exceed the overfishing catch limits (OFL) (Prager and Shertzer, 2010; Shertzer et al., 2010). They calculate an annual acceptable biological catch (ABC) that takes into account scientific uncertainty in the estimate of OFL and an annual catch target (ACT) that further accounts for management uncertainty in controlling catch to the ABC, such that there is a pre-specified probability, P^*, that attempting to catch that target would in fact result in the fishing mortality rate exceeding the overfishing rate. The flow of an ADMB estimation program does not allow for iteratively searching for target catch levels that would satisfy this probability condition because the uncertainty is not known until the program completes. However, SS takes a multiple pass approach (Appendix, Table A.2) that obtains a nearly comparable result.

The multiple passes in the forecast are needed to mimic the actual sequence of assessment to management action to catch over a multi-year period. The first pass calculates the OFL based on

catching the OFL each year, so presents the absolute maximum upper limit to catches. The second pass forecasts a catch based on a harvest policy, then applies catch caps and allocations, then updates the *F*'s to match these catches. In the third pass, stochastic recruitment and catch implementation error are implemented and the *F* that would be needed to catch the adjusted catch amount previously calculated in the second pass determined. With this approach, SS is able to produce estimates of the probability that *F* would exceed the overfishing *F*, F_{lim}. In effect, it is the complement of the P^* approach. The P^* approach calculates the future stream of annual catches that would have a specified annual probability of $F > F_{\text{lim}}$, while SS calculates the expected time series of probabilities that the *F* resulting from a designated harvest policy would exceed a specified level.

3.8. Penalized log likelihood

Some data are principally influential in estimating the growth of individuals, others in estimating trends in population abundance, others in the level of fishing mortality, and others in fluctuations in recruitment to the stock. The interactions among these various population and observational factors can be investigated and their combined contribution to the variance in assessment results calculated by including all data in a common framework and using a single overall weighted penalized log-likelihood. Calculation of the log-likelihood for each datum requires an observation error specification to scale all deviations in equal terms.

The error distribution that is typically used for abundance data is the lognormal, although the normal and Student's *t*-distributions are also available. The Student's *t*-distribution, with a user-specified low number of degrees of freedom, provides a tool to reduce the influence of outliers (Chen et al., 2000).

The multinomial is used for composition data because it inherently assigns more emphasis on fitting data for the composition bins that have higher expected proportions. The multinomial uses the assigned sample size to scale the variance of each bin in a composition vector. This sample size is typically in the range of 50–200 individuals, which is much lower than the actual number of individuals measured because of correlation within the sampling process (Crone and Sampson, 1998). SS allows for a constant to be added to each element of the observed and estimated composition vectors, thus providing increased robustness when data are noisy. SS also allows for the tails of each vector to be accumulated until there is a user-specified proportion in the terminal element of the vector. This tail compression feature causes the model to not make likelihood calculations from a large number of vector elements in the long tail, each of which is near zero, highly auto-correlated, and with low precision. Instead, the accumulated tail will have a higher proportion and thus will get more influence in model fitting. The tail compression feature should be used judiciously when data are sparse because an excessive degree of bin accumulation could occur for composition vectors that are based on just a few fish. A model output based on the multinomial distribution is the "effective sample size" (McAllister and Ianelli, 1997). This value represents the size of the random sample needed, on average, to achieve a fit that is as good as the variance in the model's fit to the composition vector.

The contribution of each datum to the overall log-likelihood function in SS is inherently weighted by the measurement variance of that datum. The user has options to increase or decrease the variances of the input data or, equivalently, the imposed weighting on their contribution to the likelihood function because these variances are rarely known exactly and because extra process error often must be dealt with as measurement error. For abundance data, SS allows for estimation of a parameter for each abundance time series to adjust the standard error of all its observations such that the root mean squared error of the model fit to that abundance time series is similar to the averaged assigned standard error of each observation. For composition data, input sample sizes should be adjusted to be at approximately the same level as the estimated effective sample sizes to achieve good variance estimates. Tools for iterative adjustment of these sample sizes are provided, but, following Maunder (2011), it may be possible to use the effective sample size calculation in a revised objective function to allow for internal estimation of a parameter representing a sample size adjustment factor.

The objective function in SS contains penalties on parameters and process deviations that are analogous to Bayesian priors, as is common in integrated analysis models (Maunder and Punt, in this volume). These penalties are based on the assumed prior variances for the parameters, and can either be designed to be informative, or be broadly uninformative. Uninformative priors serve to stabilize model performance when data are uninformative about some of the parameters. Examples of informative priors include those for each annual deviation in recruitment being penalized according to the assumed or estimated variability among all recruitment deviations, those for the parameters determining the random walks in a parameter changing over time, and those for parameters such as natural mortality. These penalties are interpreted as priors when the MCMC algorithm is used to sample parameters vectors.

Correct relative weighting among likelihood components is crucial to attain good model performance and good estimates of the variances of the model results. The model result is a complex weighting of the information in the data and the penalties (Methot and Taylor, 2011). For example, when multiple CPUE time series are used as indicators of stock abundance, the final result is a weighted average of all their trends. If a particular fleet's CPUE trend does not match those for the other fleets, its contribution to the overall result is not diminished unless the model is configured to allow for estimation of an extra observation, or process, variance parameter to account for the needed degree of downweighting. It is situation-dependent whether the adjustment should be made on the basis of extra observation variance, which makes the poor fit more likely, or on the basis of extra process variance, which creates a phantom process to bring the ill-fitting data into better conformance with other information in the model. However, the result may not stabilize at an intermediate solution when strongly contradictory information are included in an assessment. Instead, alternative local minima in the likelihood surface may be created and the estimation procedure could be trapped in a local minima and fail to identify the global minima. Patterns in residuals should be scrutinized for evidence that the model configuration has insufficient flexibility to reduce this pattern to white noise consistent with the variance of the data source.

3.9. Variance estimation

SS employs several ways to represent uncertainty in the results. ADMB uses the quadratic curvature of the multi-dimensional log-likelihood surface around the maximum likelihood set of parameters to calculate a variance-covariance matrix (Fournier et al., 2011), then uses a Taylor series approximation to propagate this measure of parameter uncertainty to the derived quantities. In SS, the derived quantities include the time series of spawning biomass, recruitment, fishing mortality index, stock depletion, and various reference points such as maximum sustainable yield. The time series of derived quantities extend into the forecast era, providing a continuous sequence of estimated derived quantities and their variances. SS can operate in Monte Carlo Markov Chain mode to provide a Bayesian description of the uncertainty associated with the parameters and derived quantities. It is possible to automate a

sequence of SS runs across a range of fixed values for one or several parameters to generate a likelihood profile or matrix of results.

SS also contains a data generation routine that can provide parametric bootstrap data sets to assist in investigation of model performance and variance estimation (e.g. Piner et al., 2011). The original data are not re-sampled. Instead a new data set is created with the same variance properties that were assumed when analyzing the original data. For each datum, the form and scale of the sampling distribution assigned to that datum (i.e. the distributional assumption of the likelihood function) are used to create a random observation given the expected value for that observation. For example, a multinomial sample of 89 fish would be drawn from the estimated length-composition of a sample that had been given an input sample size of 89. Importantly, the observations in each of these new data sets will not contain any patterns in their residuals, except by chance. Consequently, any autocorrelation or unlikely pattern in the residuals of the original data relative to the model fit will not be propagated into the bootstrap data sets. It is therefore important to ensure that the assigned variance of the input data are approximately the same as the variability between the observed and expected values before creating bootstrap data sets. These bootstrap data sets contain measurement error only and do not randomize any of the process (e.g. annual fluctuations in recruitment or time-varying selectivity) error that may be a feature of the particular model setup being analyzed. Some researchers (e.g. Lee et al., 2011) have used this parametric bootstrap data generation process combined with additional code to allow for process error in processes such as recruitment or selectivity.

It is possible to obtain additional information about the robustness of the model's fit to the original data using this bootstrap approach,. When there is pattern in the residuals to the original fit, the model will be in a tension between trying to reduce those residuals while being restrained from doing so by the fit to all the other data in the model and the model structure itself. This tension does not exist among randomly generated observations and the parameter estimates will then tend to drift away from the values obtained when fitting the original data and to a new set of values consistent with the independently distributed data. Following this logic, Lee et al. (2011) and Piner et al. (2011) investigated the robustness with which natural mortality rate could be estimated by SS. Methot and Taylor (2011) used this approach to investigate model performance with regard to estimation of recruitment deviations.

Variance results from parametric, MCMC, and bootstrapping will be quite similar when the normal approximation is good and the residuals in the original model fit are sufficiently random (Figs. 2 and 3). Stewart et al. (in this volume) provides an extensive comparison of these approaches to variance estimation.

3.10. Output processing

SS, and all models that deal with complex and voluminous data, needed auxiliary tools to visualize model outputs and quickly produce report-ready tables and figures. This done in SS by providing some output in text tables delimited by keywords and some in list format that provides a filterable/sortable database. This approach allows for parsing by different systems. One is the SS graphical interface developed by Alan Seaver and found in the NMFS Assessment Toolbox (NOAA Toolbox, 2011). A second is an Excel spreadsheet distributed with SS. Third is an R-based module, termed r4ss (Taylor et al., 2011). All three work by searching for keywords, then parsing a table or list of values adjacent to that keyword. This approach provides high flexibility with minimal programming. Users are able to quickly examine results for problematic residual patterns then proceed to produce report quality graphics using these data display tools. While the Toolbox and R modules have successfully co-evolved with SS, future versions of SS could make the interaction between the input, execution and output modules more intuitive and efficient.

4. Discussion

Sixty-one stocks worldwide have been assessed using SS to date (Appendix B), and exploratory applications for many others are underway. This success is due to the flexible scalability to a variety of data and life history situations; to the improved efficiency and communication achieved by use of a consistent, standardization approach; and to the continual evolution to meet the needs of the international fishery assessment community.

4.1. Scalable to data availability

The ability of SS to flexibly and simultaneously deal with many types of data is a powerful aspect of the framework. SS is essentially completely scalable between statistical catch-at-age analysis on one extreme, and age-structured, biomass dynamics models and catch curve analysis on the other extreme. SS can estimate the selectivity characteristics of each well-sampled fishing fleet and the time sequence of recruitment to the stock, just like other statistical catch-at-age models when highly informative catch-at-age are available. SS can still estimate selectivity and recruitment, although with less precision, with more reliance on priors and with less signal being detected when composition data are sparse or represented only by length composition. SS can use a fishery selectivity asserted by the user, and collapse the estimated catch-at-age into an expected value for the overall catch and CPUE, thus behaving exactly like an age-structured production model when there is only a bulk indicator of the trend in stock abundance, for example the catch per unit effort (CPUE) of a fishing fleet. If all that exists is a time series of total catch and expert judgment on the degree of stock depletion over time, as in the depletion-based stock reduction analysis (Dick and MacCall, 2011), then SS can use this depletion estimate as a "survey" of spawning biomass at the end of the time series relative to the spawning biomass before fishing started (Cope, 2012). If the only datum was an estimate of the size composition of the catch at one point in time, SS could generate an expected value for that observation while estimating growth parameters and mortality to essentially conduct a catch curve analysis. However, SS could also directly take any information about selectivity and could analyze multiple such observations scattered over the history of the fishery into account, unlike a simple catch curve analysis.

The scalability of SS is facilitated by its definition of annual recruitment as a deviation from the expected recruitment given the spawner–recruitment relationship. All recruitments in a time series can be estimated because the deviations are penalized according to the assumed degree of variability among the recruitments. Thus, SS is essentially a stochastic stock reduction analysis (Walters et al., 2006). Given this approach, data-poor applications end up with near-zero maximum likelihood estimates of all the recruitment deviations, so all the productivity information is just in the two or three parameters of the spawner–recruitment curve. However, the recruitment deviations contribute to the variance to the final result and can take on non-zero posterior values when informative data are included in the assessment. SS partitions the assumed total variability in recruitment into a component that is among the recruitment deviations and a component that is the residual variance of each recruitment deviation (Methot and Taylor, 2011). This feature allows SS to span data-poor and data-rich eras within a single analysis. Thus, SS is scalable to the range of data quality within

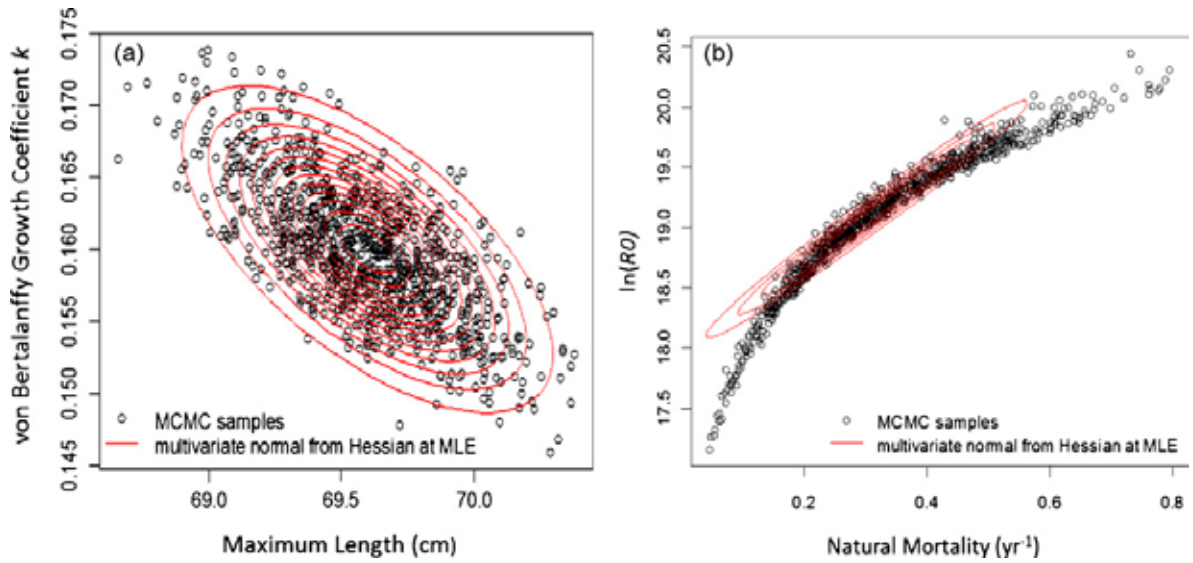

Fig. 2. Bivariate comparison of the estimated growth parameters maximum length (L_∞) and k (a). Ellipses represent multivariate normal confidence regions calculated around the maximum likelihood best estimate. The points are drawn from a converged MCMC chain. Two parameters, natural mortality and asymptotic recruitment, for which the normal approximation does not represent the tails well (b).

the time series being analyzed, as well as between applications with different data availability.

The value in using a complete model such as SS to analyze data-weak situations is that it is not necessary to switch modeling approaches as more types of data become available. Further, it forces attention on each aspect of the biologically structured population (natural mortality, growth, reproduction, selectivity, etc.) rather than collapsing these processes into a two-parameter black box as in biomass dynamics models. An informative situation was investigated during the 2012 review of assessment models for various flatfishes (ICES, 2012). The assessment had previously been conducted using ASPIC (Prager, 1994), which provides a biomass dynamics approach, with no process error, for analyzing time series of catch and indexes of stock abundance. Alternate assessment models being considered included SS and a biomass dynamics model with process error. SS was used to demonstrate how the biomass dynamics model with unconstrained process error could drift into a high productivity, high process error scenario that fit the abundance index too well. With SS, this phenomenon could be replicated, and then controlled for by using sufficient length-composition data to stabilize the result to a better compromise between process and measurement error.

4.2. Model complexity

The complexity of integrated analysis models and their ability to simultaneously analyze several diverse types of data means that investigation and control of their performance is challenging. Without informative data, unconstrained model complexity will always lead to highly uncertain results. Simpler models fit to good data can outperform more complex models. However, simple data do not necessarily imply good data nor does a simple model necessarily imply an accurate model of the relevant processes. Good data must accurately represent the state variable (fish abundance) with good fidelity, and the contrast in these data over time must be sufficient to provide information on the dynamics of the system. If data that meet these criteria are available, which is rare, a model that is just complex enough to represent the main dynamic processes should be able to track these data and

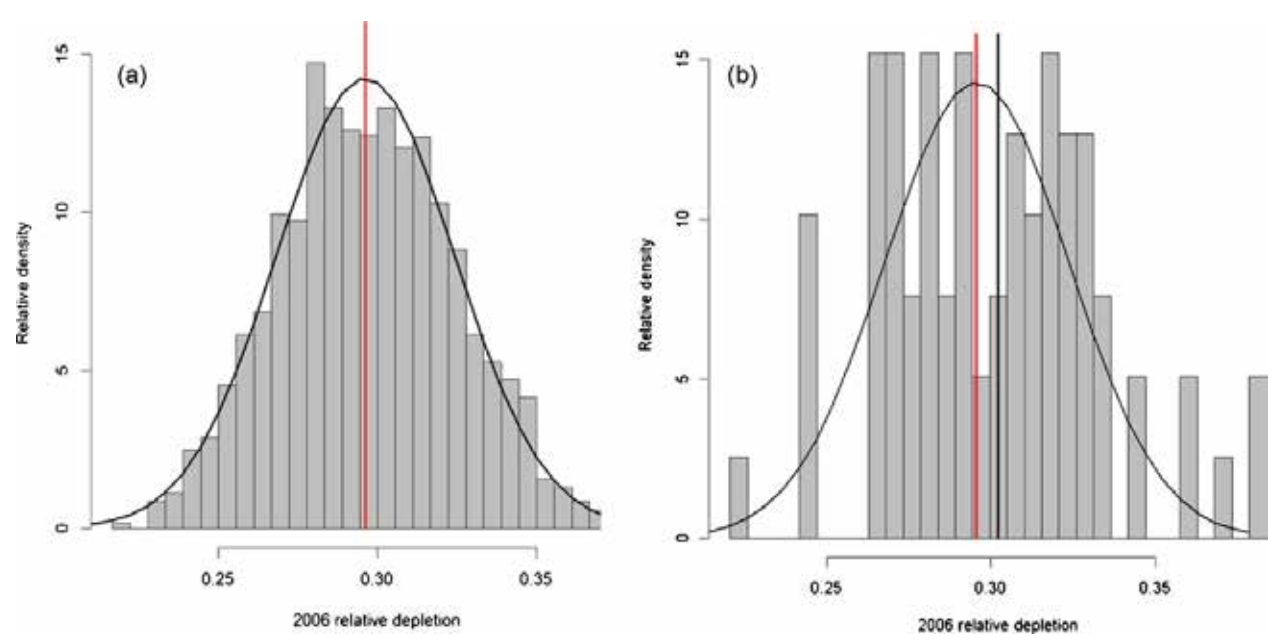

Fig. 3. The MCMC distribution of relative depletion from the 2006 Pacific whiting assessment (Helser et al., 2006) is a nearly exact match for the parametric estimate of variance in this quantity (a). The right panel (b) shows good comparison between the normal error in this quantity and the results of 75 bootstrap runs. Dark vertical line is the median of the bootstrap estimates.

Please cite this article in press as: Methot Jr., R.D., Wetzel, C.R., Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. (2012), http://dx.doi.org/10.1016/j.fishres.2012.10.012

provide reasonable inference about the population. However, difficulties in measuring fish abundance mean that the time series of abundance indices commonly available for stock assessments are typically contaminated by other processes (dome-shaped selectivity, time-varying selectivity, density-dependent catchability, etc.), and the ecosystem processes affecting fish populations are more complex than a two-parameter parabola. More complex models are needed to take a structured approach to investigate and then account for the many factors that have created patterns in the available data. SS provides this biologically structured approach in the form of size-selectivity, aging imprecision, retention functions, flexible approaches to time-varying parameters, and other factors designed to extract information on fish stocks from a diversity of data types.

SS and all integrated analysis models must be configured to achieve a good balance between data and model complexity. Results are based on the weighted influence of all the data included in the model. The model does not ignore data that do not fit well. It tries to fit each datum in accord with the measurement error associated with that data type, and it uses all free parameters to attempt to improve the goodness of fit. When the data contain patterns that are more complex than can be produced by the model, it will see this complexity as observation deviations from its too simple structure. It will then tend to inflate its estimate of the observation error associated with these data to resolve the lack of fit, and/or it will use other, inappropriate, parameters to attempt to fit the data. For example, consider a model that is presented with complete catch-at-age data that comes from a fishery with a trend in selectivity over time, but the model is not configured to allow for time-varying selectivity. Two things will happen: the time series of estimated recruitments will be biased by trying to mimic time-varying selectivity, and there may be an inflated variance estimate for the age-composition data because recruitment trends cannot fully capture time-varying selectivity. On the other hand, attempting to estimate unconstrained time-varying fishery selectivity in a situation with weak or no age-/size-composition data will result in broad confidence intervals, if model convergence can even be attained in this over-parameterized situation. Model complexity must be appropriate given the available data. One way to accomplish this is to allow the model to have substantial flexibility, mostly as time-varying process error in various parameters, but to also include informative priors on the extent of variation to stabilize estimation when data are not informative. Specification of these informative priors is one of the greatest challenges for good practices in the application of complex fishery assessment models.

Evaluation of the appropriate degree of model complexity is not easy. For example, what are the implications of allowing for dome-shaped selection patterns that change over time, ignoring aging imprecision, or allowing the model to estimate natural mortality rates? Each of these, and many other, questions can be investigated by using one model to generate data with known characteristics from a simulated population, and another model to attempt to estimate that population using the generated data. Ideally, the data generating system and the estimation model would be implemented independently to help assure a robust test of the issues being investigated. He et al. (2011) and Helu et al. (2000) used such an approach to investigate the ability of SS to estimate selectivity. However, consistency between the data generation and estimation models can help ensure that performance is due to the ability of the estimation model to extract information from the data and not due to subtle differences between the generation and the estimation models when the issues are mostly about the ability of the estimation model to derive information from particular types of data or to operate in situations with limited data contrast. The parametric bootstrap feature in SS provides a tool to generate the needed data sets.

4.3. Linkage to the ecosystem and the environment

An important feature of SS is its ability to serve as a bridge to environmental and ecosystem investigations. The earliest application of SS (Methot, 1986) included the influence of a predator on natural mortality as well as a temperature-effect on maturation, which led to a time-varying linkage between the adult stock and the expected larval abundance survey. Predators can be included in SS because they effectively operate as an additional fishing fleet. If the predator's total annual consumption is known, then this is the catch. If the predator abundance time series is known, then this serves as an index of the predator's "effort". If the ages or sizes of the prey are available from stomach samples, then this serves as the size-/age-composition of the predator's catch. Livingston and Methot (1998) took such an approach in investigating the role of several predators on walleye pollock *Theragra chalcogramma* off Alaska.

The time series of annual fluctuations in recruitment is one of the key outputs from assessment models and the possibility of including environmental covariates in assessment models has attracted much attention in the literature (Maunder and Watters, 2003; Haltuch and Punt, 2011). Where sufficient age- or size-composition data exist, age-structured models such as SS are able to estimate the time series of recruitment empirically. However, estimation of recruitment fluctuations degrade when data quality is poor, as occurs during historical periods and certainly as one extends into the forecast period (Maunder et al., 2006; Methot and Taylor, 2011). In some cases, environmental correlates of recruitment have been surmised and these environmental factors, such as ocean temperature or ocean winds, are often available further back in time than age-structured fishery data, and are available with only short time lag up to present. Environmental correlates of recruitment fluctuations can be included in SS in two ways (Schirripa et al., 2009). In one approach, the environmental time series drives the expected recruitment deviation through an estimated link parameter. Most parameters in SS can have such an environmentally driven link. In the other approach, the environmental information is treated as data relevant to the expected magnitude of the recruitment deviation, rather than a causal factor for the deviation. This approach is more robust when there are missing values in the environmental time series. Brandon et al. (2007) used such an approach in a model of the gray whale *Eschrichtius robustus* population off the west coast of North America. Conceptually and algebraically, this is tantamount to asserting that measuring ocean temperature (or some other relevant environmental factor) is equivalent to conducting a survey of the survival rate of pre-recruit juveniles. In both cases, the information is interpreted to be an index, with measurement error, of the actual recruitment. With small modification, SS could evolve to include this "environment as data" approach for more model processes that exhibit change over time.

4.4. Benefits of standardized models

A comprehensive modeling framework such as SS enhances communication, efficiency, and education in the fishery assessment community (Methot, 2009). Communication is enhanced by creating a familiarity among users, reviewers, and clients regarding terminology and approach. Reviewers who are already familiar with SS can quickly focus on key issues for the assessment being reviewed, rather than spend time learning the features of a novel assessment model. Imagine an assessment workshop with three assessments each with a different modeling system and each with discards treated in a different manner. Communication chaos will

ensue and detract from the effectiveness of the workshop. Standardized models also enhance communication and efficiency by stimulating development of a mature set of tools, such as the output processing package r4ss. These standardized approaches and products will reduce the time it takes to produce an assessment report.

Students and stock assessment training can also be beneficiaries of standardized assessment approaches. Numerous students at School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA (e.g. Wetzel and Punt, 2011; Garrison et al., 2011) and elsewhere have honed their programming and population dynamics skills by developing simulation models to test the performance of SS in various situations, as well as to be involved in stock assessments that use SS (Stewart et al., 2011).

The challenge for the fishery assessment, research and education community is to find the right balance between assessment model standardization, which enhances communication and efficiency, and assessment model development, which provides for highly focused investigation and enhances research opportunities and education. Achieving the benefits of both standardization and innovation will require more discipline from the community to use only a core set of models to produce assessment results and to refrain from conducting research as part of assessments designed to inform management. The solution also needs to encourage research into better assessment approaches and more complete testing of model performance to achieve a set of good practices for application of the standard models. The third step would be creation of a pathway by which the results of the research will periodically be reviewed and integrated into the standardized models. A suite of standardized models are commonly employed in climate and oceanographic settings. It seems advisable for the fish assessment community to move in this direction to achieve improved throughput and communication.

4.5. Future evolution of SS

The features of SS have never been static for long and continue to evolve in response to the needs and ideas of the fishery assessment community. All features are contained within one ADMB file with approximately 20,000 lines of code; there has been no splintering into divergent pathways to meet individual needs. While it might seem that this large internal structure containing numerous arrays dimensioned at runtime would result in slow execution, in fact the speed of execution depends principally on the complexity of the user-defined set-up and the amount of the data. A simple age-structured production model will run to completion in a few seconds, while a multi-area model containing rich size-composition data from numerous fleets will take several seconds for each iteration of the estimation procedure and many hours to converge. Undoubtedly, customized models for a particular assessment will execute faster than SS on the same problem, but the overhead to attain full flexibility is small.

A recap of some of the features introduced into SS in 2011 illustrates the pace of evolution.

- a new survival-based spawner–recruitment relationship was used for dogfish *Squalus suckleyi* (Taylor et al., in this volume);
- the degree of depletion of the spawning stock can now be used (Cope, 2012) in the same manner as in Depletion-Corrected Stock Reduction Analysis (Dick and MacCall, 2011);
- a cubic spline selectivity function was introduced to allow more flexibility and potentially produce multi-modal patterns of size-selectivity. This feature was used during the 2011 assessment of west coast sablefish *Anoplopoma fimbria* (Stewart et al., 2011).
- age-specific growth coefficients were introduced to allow evaluation of unexpected patterns in mean size-at-age, as observed in some tuna species, and comparison to results explainable by size-selectivity alone;
- the capability to create an aging matrix from estimable model parameters was introduced in recognition of the growing interest in accounting for aging imprecision and bias (Cope and Punt, 2007), and was used in the assessment of Pacific cod *Gadus macrocephalus* (Thompson et al., 2011).

Evolution will continue to occur to incorporate new research findings. In some cases, SS itself has been the platform by which the new concept has been investigated. For example, Piner et al. (2011) used the parametric bootstrap feature of SS to provide an approach for investigating robustness of model results. Methot and Taylor (2011) used SS to investigate the relationship between data quality, recruitment estimation, and bias in time series. These findings were made through use of SS, but each is applicable to all integrated analysis type models.

Many ideas are under development. Some of these are:

- alternative likelihood functions for composition data (Maunder, 2011) that would allow for estimation of the variance;
- area-specific spawner–recruitment relationships, rather than the current global relationship;
- explicit elapsed time between spawning and recruitment to more closely match fish life history;
- priors on derived quantities such as selectivity;
- use of environmental time series as indicators of time-varying processes;
- internal estimation of process error for time-varying processes;
- a smoothing algorithm in the forecast so that the target catch one year cannot change too rapidly from the previous year's catch, as commonly is implemented in management procedure (Butterworth and Punt, 1999).

While addition of new features is the most common request from SS users, another is the need for practices to use the flexible features of SS and other integrated analysis models. Too often, a new user starts by invoking more SS features than are supportable by their data and by not taking a structured approach that starts from a simple set-up and gradually adds complexity and flexibility to investigate the phenomena that created their data. User guides to assist in best practices are needed as SS expands to be used by more broadly.

SS, in common with nearly all fishery assessment models, is the product of a single researcher who started out to build an analytical tool for a particular application. While such an approach is common and well-suited to scientific investigations, the importance of the results of stock assessment for government regulatory processes would seem to justify a more concerted effort. As the ADMB-based version of SS reaches its 10 year mark in 2012, it seems time to pause and take stock of alternative development pathways. One is status quo is which gradual evolution of the code is undertaken and additional scientists slowly get engaged in more aspects of code development. Another would be to embark on a more concerted effort to assure that a successor model would even more completely meet the needs of the assessment community. This effort could involve a broadly based steering group to provide guidance for model development, and professional programmers to turn their ideas into well-written computer code. It would take a serious look at the pathway from raw data to model inputs so that reconfiguration of model inputs is streamlined and not error-prone. It would co-develop the model engine with the user interface so that control of the model is intuitive and straightforward. The model engine itself would need a strong flexible framework and modular implementation to ease evolution of new features. Assembling the funding and infrastructure to implement and maintain such a

collaborative model will not be easy. However, highly publicized controversies over assessment results highlight the importance of putting sufficient effort into building the best tools possible to take advantage of the great wealth of fishery and survey data being collected today.

Acknowledgments

Stock synthesis is the career-spanning result of the efforts of Richard Methot, but it would not have succeeded without the strong support, encouragement and input of numerous colleagues. Dave Fournier led the way with his pioneering work in the 1980s and subsequent development of ADMB. NOAA colleagues Jim Ianelli, Larry Jacobson, Alec MacCall, Michael Schirripa, Ian Stewart, Ian Taylor, Grant Thompson, and others have explored and tested model features in their assessments and have been instrumental in providing ideas and insight for continued model development. Mark Maunder and André Punt have been steadfast supporters and advocates for the SS approach. Students, particularly at University of Washington, are also an integral part of the SS village as they learned assessments through its application, and developed as scientists through testing of SS features in their studies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2012.10.012.

References

Aires-da-Silva, A., Maunder, M.N., 2012. Status of bigeye tuna in the eastern Pacific Ocean in 2011 and outlook for the future. Inter-American Tropical Tuna Commission, Stock Assessment Report 13, pp. 18–29.
Branch, T.A., 2009. Differences in predicted catch composition between two widely used catch equation formulations. Can. J. Fish. Aquat. Sci. 66, 126–132.
Brandon, J.R., Punt, A.E., Wade, P.R., Perryman, W.L., Methot, R.D., Maunder, M.N., 2007. Incorporating environmental time series into a population dynamics model for eastern North Pacific gray whales. IWC Scientific Paper SC/59/BRG26.
Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., 2005. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v2.07-2005/08/21. NIWA Technical Report 127.
Butterworth, D.S., Punt, A.E., 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. 56, 985–998.
Butterworth, D.S., Ianelli, J.N., Hilborn, R., 2003. A statistical model for stock assessment of southern Bluefin tuna with temporal changes in selectivity. Afr. J. Mar. Sci. 25, 331–361.
Chen, Y., Breen, P.A., Andrew, N.L., 2000. Impacts of outliers and mis-specification of priors on Bayesian fisheries-stock assessment. Can. J. Fish. Aquat. Sci. 57, 2293–2305.
Cope, J.M., 2012. Implementing a statistical catch-at-age model (stock synthesis) as a tool for deriving overfishing limits in data-limited situations. Fish. Res., http://dx.doi.org/10.1016/j.fishres.2012.03.006.
Cope, J.M., Punt, A.E., 2007. Admitting ageing error when fitting growth curves: an example using the von Bertalanffy growth function with random effects. Can. J. Fish. Aquat. Sci. 64, 205–218.
Crone, P.R., Sampson, D.B., 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In: Funk, F., Quinn, II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks, pp. 355–370.
Dick, E.J., MacCall, A.D., 2011. Depletion-based stock reduction analysis: a catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110, 331–341.
Fournier, D., Archibald, C.P., 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39, 1195–1207.
Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. Can. J. Fish. Aquat. Sci. 55, 2105–2116.
Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2011. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 233–252.
Garrison, T.M., Hamel, O.S., Punt, A.E., 2011. Can data collected from marine protected areas improve estimates of life-history parameters? Can. J. Fish. Aquat. Sci. 8, 1761–1777.
Goodyear, C.P., 1993. Spawning stock biomass per recruit in fisheries management: foundation and current use. In: Smith, S.J., Hunt, J.J., Rivard, D. (Eds.), Risk Evaluation and Biological Reference Points for Fisheries Management. Can. Spec. Publ. Fish. Aquat. Sci. 120, 67–81.
Goodyear, C.P., 1997. Fish age determined from length: an evaluation of three methods using simulated red snapper data. Fish. Bull. 95, 39–46.
Haltuch, M.A., Punt, A.E., 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Can. J. Fish. Aquat. Sci. 68, 912–926.
He, X., Ralston, S., MacCall, A.D., 2011. Interactions of age-dependent mortality and selectivity functions in age-based stock assessment models. Fish. Bull. 109, 198–216.
Helser, T.E., Stewart, I.J., Fleischer, G.W., Martell, S., 2006. Stock Assessment of Pacific Hake (Whiting) in U.S. and Canadian Waters in 2006. Pacific Fishery Management Council, Portland, OR www.pcouncil.org
Helu, S.L., Sampson, D.B., Yin, Y., 2000. Application of statistical model selection criteria to the stock synthesis assessment program. Can. J. Fish. Aquat. Sci. 57, 1784–1793.
International Council for the Exploration of the Sea (ICES), 2012. Report of the Benchmark Workshop on Flat-fish Species and Anglerfish (WKFLAT). ICES Advisory Committee, Bilbao, Spain, 283 pp.
Kell, L.T., Ortiz, M., 2011. A comparison of statistical age estimation and age slicing for Atlantic Bluefin tuna (Thunnus thynnus). Collect. Vol. Sci. Pap. ICCAT 66, 948–955.
Laurec, A., Shepherd, J.G., 1983. On the analysis of catch and effort data. J. Cons. Int. Explor. Mer. 41, 81–84.
Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fish. Res. 109, 89–94.
Livingston, P.A., Methot, R.D., 1998. Incorporation of predation into a population assessment model of eastern Bering Sea walleye pollock. In: Funk, F., Quinn, II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, pp. 663–678.
Lo, N.C.H., Hunter, J.R., Moser, H.G., Smith, P.E., Methot, R.D., 1992. The daily fecundity reduction method: a new procedure for estimating adult fish biomass. ICES J. Mar. Sci. 49, 209–215.
Mace, P.M., Sissenwine, M.P., 1993. How much spawning per recruit is necessary? In: Smith, S.J., Hunt, J.J., Rivard, D. (Eds.), Risk Evaluation and Biological Reference Points for Fisheries Management. Can. Spec. Publ. Fish. Aquat. Sci. 120, 101–118.
Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fish. Res. 109, 311–319.
Maunder, M.N., Punt, A.E. A review of integrated analysis in fisheries stock assessment. Fish. Res., in this volume
Maunder, M.N., Watters, G.M., 2003. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and examples. Fish. Bull. 101, 89–99.
Maunder, M.N., Harley, S.J., Hampton, J., 2006. Including parameter uncertainty in forward projections of computationally intensive statistical population dynamic models. ICES J. Mar. Sci. 63, 969–979.
McAllister, M.K., Ianelli, J.N., 1997. Bayesian stock assessment using catch-age data and the sampling – importance algorithm. Can. J. Fish. Aquat. Sci. 54, 284–300.
Methot, R.D., 1986. Synthetic estimates of historical abundance and mortality for northern anchovy, *Engraulis mordax*. NMFS Southwest Fisheries Science Center Admin. Rep. 12-86-29.
Methot, R.D., 1989. Synthetic estimates of historical and current biomass of northern anchovy, *Engraulis mordax*. Am. Fish. Soc. Sympos. 6, 66–82.
Methot, R.D., 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Int. N. Pac. Fish. Comm. Bull. 50, 259–277.
Methot, R.D., 2000. Technical description of the Stock Synthesis assessment program. NOAA Tech Memo. NMFS-NWFSC-43.
Methot, R.D., 2009. Stock assessment: operational models in support of fisheries management. In: Beamish, R.J., Rothschild, B.J. (Eds.), Future of Fishery Science – Proceedings of the 50th Anniversary Symposium of the American Institute of Fishery Research Biologists. Fish & Fisheries Series, vol. 31. Springer, Seattle, WA, pp. 137–165.
Methot, R.D., Dorn, M.W., 1995. Biology and fisheries for North Pacific hake (*M. productus*). In: Alheit, J., Pitcher, T.J. (Eds.), Hake: Fisheries, Ecology and Markets. Chapman & Hall, London, pp. 389–414.
Methot Jr., R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68, 1744–1760.
Nichols, S., 1986. Updated yield-per-recruit information on the Tortugas pink shrimp fishery. N. Am. J. Fish. Manage. 6, 339–343.
NOAA Fisheries Toolbox, 2011. Stock Synthesis, Version 3.23b. http://nft.nefsc.noaa.gov
Parrish, R.H., Mallicoate, D.L., Klingbeil, R., 1986. Age dependent fecundity, number of spawnings per year, sex ratio, and maturation stages in northern anchovy, *Engraulis mordax*. Fish. Bull. 84, 503–517.
Piner, K.R., Lee, H.-H., Maunder, M.N., 2011. A simulation-based method to determine model misspecification: examples using natural mortality and population dynamics models. Mar. Coast. Fish. 3, 336–343.
Pope, J.G., 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. Int. Comm. Northwest Atl. Fish. Res. Bull. 9, 65–74.

Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92, 374–389.
Prager, M.H., Shertzer, K.W., 2010. Deriving acceptable biological catch from the overfishing limit: implications for assessment models. N. Am. J. Fish. Manage. 30, 289–294.
Sampson, D.B., Scott, R.D., 2012. An exploration of the shapes and stability of population-selection curves. Fish. Fish. 13, 89–104.
Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-Am. Trop. Tuna Comm. 1, 27–56.
Schirripa, M.J., Goodyear, C.P., Methot, R.M., 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. ICES J. Mar. Sci. 66, 1605–1613.
Schnute, J.T., 1981. A versatile growth model with statistically stable parameters. Can. J. Fish. Aquat. Sci. 38, 1128–1140.
Shertzer, K.W., Prager, M.H., Williams, E.H., 2010. Probabilistic approaches to setting acceptable biological catch and annual catch targets for multiple years: reconciling methodology with national standards guidelines. Mar. Coast. Fish. 2, 451–458.
Stewart, I.J., Thorson, J.T., Wetzell, C., 2011. Status of the U.S. Sablefish Resource in 2011. Pacific Fishery Management Council, Portland, OR www.pcouncil.org
Stewart, I.J., Hicks, A.C., Taylor, I.G., Thorson, J.T., Wetzel, C., Kupschusc, S. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fish. Res., in this volume
Taylor, I.G., Methot, R.D. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res., in this volume
Taylor, I.G., Stewart, I.J., Hicks, A., Garrison, T.M., Punt, A.E., Wallace, J.R., Wetzel, C.R., 2011. r4ss: R code for Stock Synthesis. R package version 1.16. http://R-Forge.R-project.org/projects/r4ss/
Taylor, I.G., Gertseva, V., Methot, R.D. Jr., Maunder, M.N. A stock-recruitment relationship based on pre-recruit survival, illustrated with application to spiny dogfish shark. Fish. Res., in this volume
Thompson, G.G., A'Mar, Z.T., Palsson, W.A., 2011. Assessment of the Pacific cod stock in the Gulf of Alaska. In: Plan Team for Groundfish Fisheries of the Gulf of Alaska (compiler), Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. North Pacific Fishery Management Council, 605 W. 4th Avenue Suite 306, Anchorage, AK 99501, pp. 161–306.
Walters, C.J., Martell, S.J.D., Korman, J., 2006. A stochastic approach to stock reduction analysis. Can. J. Fish. Aquat. Sci. 63, 212–223.
Wetzel, C.R., Punt, A.E., 2011. Performance of a fisheries catch-at-age model (stock synthesis) in data-limited situations. Mar. Freshw. Res. 62, 927–936.

Please cite this article in press as: Methot Jr., R.D., Wetzel, C.R., Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. (2012), http://dx.doi.org/10.1016/j.fishres.2012.10.012

Conserv Genet
DOI 10.1007/s10592-012-0432-0

RESEARCH ARTICLE

Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure

Kristen Ruegg · Howard C. Rosenbaum · Eric C. Anderson · Marcia Engel · Anna Rothschild · C. Scott Baker · Stephen R. Palumbi

Received: 11 May 2012 / Accepted: 23 November 2012
© Springer Science+Business Media Dordrecht 2012

Abstract Once hunted to the brink of extinction, humpback whales (*Megaptera novaeangliae*) in the North Atlantic have recently been increasing in numbers. However, uncertain information on past abundance makes it difficult to assess the extent of the recovery in this species. While estimates of pre-exploitation abundance based upon catch data suggest the population might be approaching pre-whaling numbers, estimates based on mtDNA genetic diversity suggest they are still only a fraction of their past abundance levels. The difference between the two estimates could be accounted for by inaccuracies in the catch record, by uncertainties surrounding the genetic estimate, or by differences in the timescale to which the two estimates apply. Here we report an estimate of long-term population size based on nuclear gene diversity. We increase the reliability of our genetic estimate by increasing the number of loci, incorporating uncertainty in each parameter and increasing sampling across the geographic range. We report an estimate of long-term population size in the North Atlantic humpback of ~112,000 individuals (95 % CI 45,000–235,000). This value is 2–3 fold higher than estimates based upon catch data. This persistent difference between estimates parallels difficulties encountered by population models in explaining the historical crash of North Atlantic humpback whales. The remaining discrepancy between genetic and catch-record values, and the failure of population models, highlights a need for continued evaluation of whale population growth and shifts over time, and continued caution about changing the conservation status of this population.

Electronic supplementary material The online version of this article (doi:10.1007/s10592-012-0432-0) contains supplementary material, which is available to authorized users.

K. Ruegg (✉) · S. R. Palumbi
Department of Biology, Hopkins Marine Station,
Stanford University, 120 Oceanview Boulevard,
Pacific Grove, CA 93950, USA
e-mail: kruegg@ucsc.edu

H. C. Rosenbaum
Ocean Giants Program, Global Conservation,
Wildlife Conservation Society, Bronx, NY, USA

H. C. Rosenbaum · A. Rothschild
Sackler Institute for Comparative Genomics,
American Museum of Natural History,
79th Street and Central Park West,
New York, NY, USA

E. C. Anderson
Fisheries Ecology Division, Southwest Fisheries Science Center,
National Marine Fisheries Service, NOAA, 110 Shaffer Road,
Santa Cruz, CA 95060, USA

E. C. Anderson
Department of Applied Math and Statistics,
University of California, Santa Cruz,
CA 95060, USA

M. Engel
Instituto Baleia Jubarte/Humpback Whale Institute,
Rua Barao do Rio Branco 26, Caravelas,
BA 45900-000, Brazil

C. S. Baker
Marine Mammal Institute, Hatfield Marine Science Center,
Oregon State University, 2030 SE Marine Science Drive,
Newport, OR 97365, USA

Published online: 13 December 2012

Keywords Effective population size · Humpback whale · Census population size · Population structure

Introduction

Over-exploitation has resulted in the collapse of many marine populations (Pauly et al. 1998; Myers and Worm 2003; Estes et al. 2006). In some cases, however, national or international protection has led to the recovery of previously threatened or endangered species (reviewed within Scott et al. 2005). Humpback whales (*Megaptera novaengliae*) in the North Atlantic were severely depleted as a result of intense hunting during the 19th and 20th centuries (Mitchell and Reeves 1983; Braham 1984; Winn and Reichley 1985) and are currently listed as 'endangered' or 'vulnerable' by various governments and international conservation organizations (Klinowska 1991). Before the International Whaling Commission (IWC) banned commercial whaling in the North Atlantic in 1955, it was estimated that this population was reduced to <1,000 individuals (Mitchell and Reeves 1983; Katona and Beard 1990). After many decades of legal protection, humpback whales have increased in numbers (Stevick et al. 2003) and recent survey estimates suggest that they may be approaching 20,000 animals (Smith and Pike 2009). Such increases in population size within the North Atlantic have led the IUCN and the US to re-evaluate their conservation status.

Assessing the recovery of previously depleted populations requires knowledge of past population sizes, but robust estimates of past abundance can be difficult to attain. Different approaches to estimating pre-whaling population sizes can lead to starkly different conclusions about the extent of recovery in the North Atlantic humpback whale (Roman and Palumbi 2003; Holt and Mitchell 2004; Punt et al. 2006). Traditionally, the IWC has relied upon population dynamic models that use a combination of information on current abundance, catch records, rates of increase, and population structure to estimate changes in population size through time (Punt et al. 2006). Recent model estimates for the North Atlantic humpback whale suggest a pre-whaling population size of between 20,000 and 46,000 individuals, depending upon the catch data used (Punt et al. 2006). Given current abundance estimates of $\sim$17,700 individuals (Smith and Pike 2009), population model-based estimates suggest that humpbacks in the North Atlantic are approaching the lower boundary of their pre-whaling numbers. Alternatively, genetic-based estimates of pre-whaling abundance use the relationship between genetic diversity (θ) and effective population size (N_e) ($\theta = 4N_e\mu$, where μ is the average mutation rate) to estimate the long-term population size of North Atlantic Humpback whales (Roman and Palumbi 2003). Genetic estimates calculated using mitochondrial DNA (mtDNA) control region sequence, suggest a pre-whaling abundance of 150,000–240,000 depending upon the mutation rate employed (Roman and Palumbi 2003; Alter and Palumbi 2009). These genetic estimates suggest that there were substantially more whales prior to whaling than previously believed.

The discrepancy between estimates of pre-whaling abundance based upon catch records and estimates based upon genetic variability has been the subject of vigorous debate (Lubick 2003; Holt and Mitchell 2004; Clapham et al. 2005). Some argue that unavoidable uncertainties in the catch record may have led to underestimates in the number of whales removed from the North Atlantic due to whaling (Palumbi and Roman 2007). However, a recent review and re-reading of whaling records revealed only slight increases in the numbers of North Atlantic humpback whales estimated to be killed as result of whaling (from 29,000 to 30,852 total catches) (IWC 2002, 2003; Smith and Reeves 2010). Others argue that genetic estimates of long-term abundance may be inaccurate as a result of reliance on a single locus, uncertainty surrounding mutation rates and generation times, the potential influence of incomplete sampling, and the evolutionary time-scale to which a genetic estimate applies (Lubick 2003; Holt and Mitchell 2004; Clapham et al. 2005). While recent estimates of long-term population size in gray and minke whales have reduced some of these uncertainties through a variety of methodological improvements (Alter et al. 2007; Alter and Palumbi 2009; Ruegg et al. 2010), humpback whales are particularly challenging because of their complex oceanic and worldwide population structure (Baker et al. 1993; Palsboll et al. 1995; Rosenbaum et al. 2009).

Population structure may affect estimates of long-term effective population size (N_e) in a variety of ways depending upon the extent of isolation between populations. Theoretical models suggest meta-population N_e and sub-population N_e converge as migration between sub-populations increases (Hudson 1991; Waples 2010). It has also been shown that even low migration rates between sub-populations can cause an estimate of long-term N_e that is based upon samples from one sub-population to approximate the long-term N_e of the whole meta-population population (Hudson 1991). Thus, if there is migration between sub-populations, then the absence of samples from one sub-population should have very little effect on an estimate of long-term N_e for the whole meta-population because the two values will be equivalent. Alternatively, if there is no migration between populations and one uses samples from an isolated sub-population, then an estimate of long-term N_e for the meta-population will be downwardly biased. Thus, in order to identify the impact of population structure on N_e it is important to also measure levels of population structure.

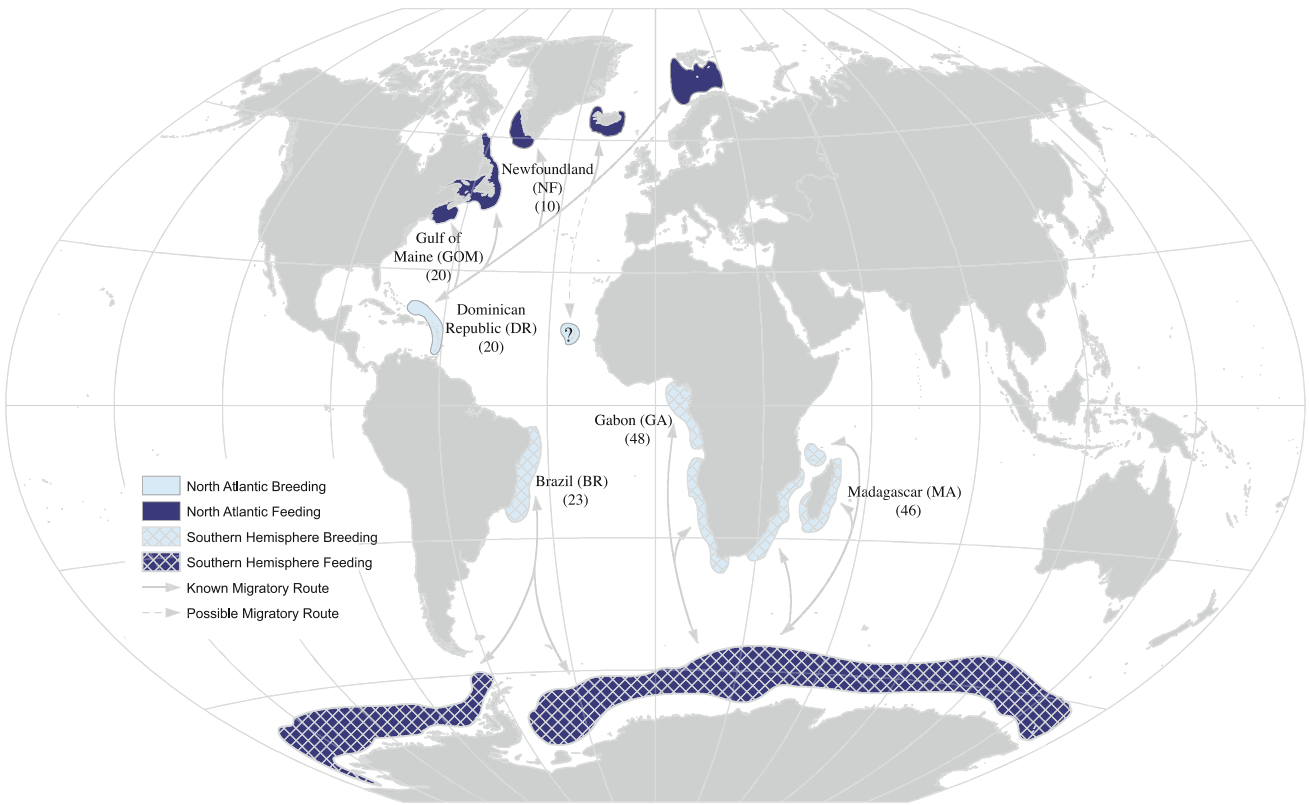

Fig. 1 Approximate breeding and feeding distributions of the North Atlantic humpack whale and 3 stocks of the Southern Hemisphere humpback whale (as described in Rosenbaum et al. 2009; Johnson and Wolman 1984). *Arrows* represent hypothesized migratory pathways. Sampling location names are followed by the number of samples in parentheses

Presently, humpback whales are divided into three oceanic populations, the North Atlantic, the North Pacific and the Southern Hemisphere, based on genetic and tagging data suggesting limited migration between ocean basins (Mackintosh 1965; Baker et al. 1993). Previous analysis of worldwide population structure based upon mtDNA suggests that humpback whales from the North Atlantic are most strongly differentiated from those in the North Pacific and less strongly differentiated from those in the Southern Hemisphere (Table 2, Baker et al. 1993). Strong divergence between North Atlantic and North Pacific humpback whales is thought to result from the fact that sea ice has likely blocked the main northern migratory corridor between the two groups since the Sangamonian Interglacial period (~140,000 years ago). As a result, genetic diversity within the North Atlantic is unlikely to be strongly influenced by past migration from the North Pacific. Thus, while we will test the assumption that gene flow with the North Pacific does not influence N_e in the North Atlantic, the main focus of our analysis will be on populations from the North Atlantic and the Southern Hemisphere.

Humpback whales within the North Atlantic and Southern Hemisphere exhibit varying degrees of within-ocean sub-population structure resulting from complex patterns of breeding, feeding, and migration specific to each ocean region (Fig. 1). North Atlantic humpback whales show site fidelity to several discrete feeding areas extending from the Gulf of Maine to the Barents Sea off the northern coast of Norway, but individuals from all known feeding areas congregate on a common breeding area in the West Indies (Katona and Beard 1990; Smith et al. 1999; Stevick et al. 1999). Despite overlap on the West Indies breeding grounds, significant population structure between eastern and western North Atlantic feeding aggregations have been identified using mtDNA (Kst ~ 0.04) (Palsboll et al. 1995) and nuclear loci (Fst ~ 0.036) (Valsecchi et al. 1997).

Patterns of migratory connectivity in the Southern Hemisphere are less well understood, but recent evidence based upon mtDNA suggests low, but significant sub-population structure between Southwestern Atlantic, Southeastern Atlantic, and Southwestern Indian Ocean groups (Breeding Stocks A, B, and C respectively; Fst range 0.0029–0.0166) with the Northern Indian Ocean (stock X) falling out as strongly differentiated from all other groups (Fst range from 0.0797 to 0.1473) (Rosenbaum et al. 2009). In this study, we will use multiple

nuclear loci and increased sampling from within each ocean basin to gain a better perspective on the impact of population structure on long-term population size in the North Atlantic.

We calculate the long-term population size of the North Atlantic humpback whale within the context of the worldwide population structure. We focus on an in-depth analysis of the North Atlantic and the Southern Hemisphere, with particular focus on the South Atlantic, because previous data indicate that these two populations are the most likely to have exchanged migrants during a time period that may impact an estimate of effective population size. We identify strongly differentiated populations within and between the North Atlantic and Southern Hemisphere using nine nuclear loci and a multi-locus genetic clustering method. We then estimate long-term population size of the North Atlantic humpback whale, while accounting for the possibility of migration with other strongly differentiated groups. Our new estimate of long-term population size in the North Atlantic is compared with previous genetic (mtDNA) and catch-based estimates in order to highlight remaining uncertainties in estimates of pre-whaling abundance and discuss important areas for future research.

Methods

Sample collection and sequencing

Genetic samples representing 173 individuals were collected from humpback whales across the Southern Hemisphere (South Atlantic and Indian Oceans) and the North Atlantic Ocean (for regional sample sizes see Fig. 1). Biopsy samples from living whales were collected with appropriate national permits using protocols approved by the American Museum of Natural History and the Oregon State University's Animal Care and Use Committees. Samples were preserved in 70 % ethanol or salt saturated 20 % dimethyl sulfoxide solution (DMSO) and later stored at −20 °C until processed. Total genomic DNA was extracted using a standard phenol/chloroform extraction method or using a DNAeasy tissue kit (Quiagen).

Nine nuclear loci were amplified and sequenced using standard PCR and sequencing protocols (Saiki et al. 1988; Palumbi 1995) and published primers (Lyons et al. 1997) (Table 1; doi:10.5061/dryad.bj506). Individuals were sequenced in both directions for 8 of 9 loci and sequences were trimmed so that only the highest quality sequences were included in the consensus. We found that the inclusion of the reverse direction for RHO lowered the overall sequence quality. Thus, in order to avoid the possibility of artificially inflating our estimate of genetic diversity by including low quality sequence in our analysis, we restricted our analysis of RHO to the 186 bp forward direction sequence. All variable sites for the 9 loci were checked by eye using Sequencher ver. 4.8 (Gene Codes Corporation). SNPs were verified through visual confirmation in forward and reverse sequences and/or in multiple individuals. SNPs that only occurred in one individual, could not be verified with reverse complement sequences, or could not be called with confidence were removed from the analysis. In order to ensure that our dataset did not contain replicate samples, we confirmed that no individual had the same sequence across all loci. Despite multiple attempts, not all individuals sequenced successfully for every locus, resulting in variation in the final sample sizes for each locus (NA mean 42, range 27–56; SH mean 101, range 80–117).

PHASE 2.1 (Stephens et al. 2001) was used to reconstruct gametic phase, defined as the original allele combination that an individual received from each of its parents, using a burn in of 10,000 iterations and a run length of 10,000 iterations. Using Arlequin ver. 3.0 (Excoffier et al. 2005) we found no significant linkage disequilibrium among loci after correcting for multiple comparisons. To determine if our sequences were evolving in a manner consistent with equilibrium and neutrality, Tajima (1989) and Fu (1997) tests were preformed using DnaSP (Rozas et al. 2003). In neutrally evolving sequences, both values will be approximately equal to zero, while balancing selection or population expansion will result in values that are significantly greater or less than zero, respectively. We also used DnaSP to calculate the minimum number of recombination events in the sample (Hudson and Kaplan 1985) and found that 3 of 9 loci showed evidence of recombination. As a result, in loci with evidence for recombination, coalescent simulations ($n = 1{,}000$) incorporating the per gene recombination parameter (R) were used to generate 95 % confidence intervals (CI) for both Tajima's D and Fu's Fs statistics.

Testing for population structure

An analysis of population structure was performed to investigate whether or not the major divisions within and between the North Atlantic and the Southern Hemisphere humpback whales remained with increased sampling in both regions. In order to properly account for migration that may impact our estimate of genetic variation (theta, θ) for the North Atlantic humpback whale, we test 3 population structure scenarios: (1) populations with no migration over recent evolutionary history (i.e. 4 N_e generations), (2) genetically distinct populations connected by very limited migration, and (3) sub-populations that may be biologically meaningful, but are exchanging migrants at a high enough rate that they cannot be distinguished using multi-locus clustering methods. To estimate long-term population size, sub-populations (scenario 3) were lumped

Table 1 Summary statistics for 9 introns sequenced in North Atlantic and Southern Hemisphere humpback whales

Intron	Seq. length	Ocean basin	N	N_S	N_H	R_m	π	Tajima's D	Fu's Fs
ACT	886	NA	27	3	7	2	0.0016	**2.224**	−0.801
		SH	90	5	10	2	0.0015	1.089	−1.706
CAT	500	NA	40	1	2	0	0.0010	1.691	2.138
		SH	96	4	6	1	0.0012	−0.190	−1.390
FGG	941	NA	51	1	2	0	0.0004	0.070	1.300
		SH	116	5	6	0	0.0006	−0.589	−1.490
ESD	598	NA	38	5	6	1	0.0025	1.041	**0.656**
		SH	107	6	11	2	0.0017	−0.037	−3.873
GBA	298	NA	56	1	2	0	0.0002	−0.809	−1.146
		SH	117	2	3	0	0.0004	−0.841	−1.662
LAC	560	NA	31	1	2	0	0.0009	1.563	1.943
		SH	81	2	3	0	0.0008	0.231	0.555
PLP	810	NA	29	1	2	0	0.0004	0.579	1.088
		SH	107	3	4	0	0.0005	−0.296	−0.468
PTH	267	NA	55	2	3	0	0.0014	−0.006	0.128
		SH	112	2	3	0	0.0013	0.067	0.348
RHO	186	NA	56	3	5	1	0.0037	0.381	−0.495
		SH	80	3	6	2	0.0054	1.440	−0.008

N number of individuals, *Ns* number of polymorphic sites, N_H number of distinct haplotypes as determined by PHASE, ver. 2.1 (Stephens et al. 2001), *Rm* minimum number of recombination events, π nucleotide diversity (Nei 1987)

* Numbers in bold refers to a significant deviation from neutral expectation before a bonferroni correction for multiple comparison ($p < 0.05$), as determined by coalescent simulations of the null distribution using DNAsp (Rozas et al. 2003)

into respective population categories (scenario 2) and populations with no possibility of migration with the North Atlantic over recent evolutionary history were considered separately (scenario 1).

Pairwise Fst within and between ocean basins at each locus as well as across all loci were calculated using the program Arlequin ver. 3.0 (Excoffier et al. 2005). A null distribution of Fst was generated through 1,000 permutations of the haplotypes between populations and the *p* value represents the proportion of permutations leading to an Fst larger than or equal to the observed value. To assess the potential for within and between ocean basin population structure within a multi-locus framework, we used the program Structure ver. 2.2 (Pritchard et al. 2000). Preliminary runs indicated that the power for assigning individuals to clusters dropped off significantly when individuals had missing data for more than 2 of the 9 loci. Therefore, individuals with missing data for more then 2 loci were removed from the multi-locus analysis in order to ensure that there was sufficient statistical power for assignment of all individuals to clusters. Structure requires unlinked markers, so the maximum a posteriori haplotypes from PHASE at each locus were recoded as alleles. We performed 3 independent runs at each K value (K = 1–5) using a burn-in period of 100,000 iterations and a run length of 500,000. The structure analysis was run using the admixture model with correlated allele frequencies with and without the location prior.

The location prior is intended to use location information to help identify more subtle population structure, without detecting structure that is not present (Hubisz et al. 2009) and we implement it here in an attempt to identify a signature of population structure in our data that may influence our subsequent estimate of q. Locations included the Gulf of Maine (GOM), Dominican Republic (DR), Gabon (GA), Brazil (BR) and Madagascar (BA) (Fig. 1, SI Table 1). Individuals from NF were grouped within the GOM location due to low sample size from NF, geographic proximity between NF and GOM and the lack of significant Fst values between NF and GOM (see "Results" section). We determined support for the number of clusters (K) by plotting the average ln [P(X|K)] of each model as a function of K and using the ad hoc ΔK statistic proposed by Evanno et al. (2005).

Estimating θ

Using our knowledge of population structure, we employed genetic models (Kuhner 2006) that estimate long-term N_e while explicitly accounting for the possibility of migration between populations deemed distinct according to multi-

locus clustering methods. We use LAMARC ver. 2.1.3 to simultaneously estimate θ while incorporating recombination and migration between ocean regions into the model. In contrast to summary statistic estimates of θ (θ_s, θ_π, etc.), LAMARC accounts for uncertainty in the data by integrating over the space of possible genealogies using a Markov chain Monte Carlo (MCMC) procedure. In order to account for uncertainty in the data resulting from unknown gametic phase and to accommodate inter-locus variation in mutation rate, we followed the methods described in Ruegg et al. (2010). In short, to account for unknown gametic phase, LAMARC was run on 15 realizations from PHASE's posterior distribution for each of 9 introns. In addition, as recommended by the LAMARC manual, we subsampled our data to restrict the input size for each LAMARC run to 20 sequences from each major population. Thus, for each of the 15 realizations from PHASE's posterior, LAMARC was run on a different subsample of 10 randomly chosen individuals from each population. The final result from these 15 LAMARC runs was obtained by catenating the summaries from all the runs following the recommendations in the LAMARC manual for "poor man's parallelization." (Initially we attempted 3 random subsamples from each phasing—45 total LAMARC runs—but this exceeded the memory available to LAMARC). To accommodate interlocus variation in mutation rate, we implemented the gamma model for mutation rate variation within a Bayesian framework using an extension of the LAMARC package known of as GUFBUL (Gamma Updating for Bayesians Using LAMARC; Ruegg et al. 2010).

Our main objective was to estimate θ in the North Atlantic while accounting for the possibility of migration with the Southern Hemisphere. To this end we used a 2-population migration model in LAMARC on the full dataset that included nine loci (Table 1). The fact that ice has blocked the main northern migratory corridor between the North Pacific and the North Atlantic since the Sangamonian Interglacial period (~140,000 years ago) makes gene flow between the two populations unlikely. However, to further investigate the possibility that genetic diversity in the North Atlantic is influenced by migration with the North Pacific, we ran a LAMARC analysis using a 3-population migration matrix on 6 of the nine loci for which we had sequence data. θ values generated using the 3-population model were compared to values calculated using the 2-population model for each of the 6 loci.

Calculating census population size from θ

The conversion of θ into effective population size (N_e) is based upon the relationship $\theta = 4N_e\mu$ where μ is the average mutation rate. To calculate an average μ for North Atlantic Humpback whales, and to estimate uncertainty surrounding our estimate, we followed the methods described in Ruegg et al. (2010). In short, we sampled with replacement from among 9 previously published individual locus mutation rates for humpback whales; 1 of the individual locus mutation rates (PLP) was from Alter et al. (2007), while the remaining 8 were taken from a Bayesian analysis of baleen whale phylogeny and fossil history (Jackson et al. 2009). For each re-sampled locus, a sample mutation rate was drawn from the posterior distribution of the estimated mutation rate or, for PLP, uniformly from the 95 % confidence intervals on the mutation rate. This was repeated 9 times for each bootstrap replicate, and we performed 100,000 bootstrap replicates. The mean μ and the variability around that mean was obtained from these bootstrap replicates. To convert μ from units of mutations per base pair per year into mutations per base pair per generation requires an estimate of the generation length. To approximate generation length we sampled uniformly from within a range of possible values for North Atlantic humpback whales of between 12 and 24 years (Chittleborough 1965; Roman and Palumbi 2003; Taylor et al. 2007). While this lower bound on generation time, taken from Chittleborough's (1965) estimate, may be low because of age-estimate inaccuracies, it is similar to the 14.5 year estimate for modern humpback whales from Taylor et al. (2007). Here we maintain the 12–24 year range in order to stay consistent with previous estimates of long-term population size in the North Atlantic humpback whale (Roman and Palumbi 2003), and discuss the implications of different generation times on estimates of N_e.

To convert N_e to census population size (N_c) requires an estimate of the ratio of mature adults to the effective number of adults (N_{mature}/N_e) and the proportion of juveniles in the population. Although N_{mature}/N_e is difficult to calculate in most natural populations, theory suggests this ratio approaches 2 in most populations with constant size (Nunney and Elam 1994). We based our estimate of N_{mature}/N_e on equation (1) in Nunney and Elam (1994): $N_e = N/(2-T^{-1})$, where T = generation length. To approximate juvenile abundance we used catch and survey data to calculate (no. of adults + juveniles)/(no. adults) (Chittleborough 1965; Roman and Palumbi 2003). To incorporate uncertainty in juvenile abundance we sampled uniformly from within a range of likely values for North Atlantic humpback whales.

Results

Tests for neutrality and equilibrium

Among the 9 nuclear introns, nucleotide diversity averaged 0.0014 (range 0.0002–0.0054), with an average of 5 haplotypes per locus (range 2–11) and an average of 43 samples from the North Atlantic and 101 samples from the Southern Hemisphere (Table 1). These values were similar

to other baleen whale species for which data are available (gray whales: range 0.0031–0.00016; Alter et al. 2007). While Tajima's D for ACT and Fu's FS for ESD were significantly different from the simulated null distribution given $p < 0.05$, neither remained significant after Bonferroni correction for multiple comparison (corrected $p = 0.05/18$ tests $= 0.003$). The results of the Tajima's D and Fu's FS tests suggest the loci are evolving in a manner consistent with neutrality and equilibrium (Table 1).

Population structure

Across the 9 loci, Fst ranged from 0 to 0.36 (SI Table 1), with 69 % (18 of 26) of the significant pairwise Fst values being between North Atlantic and Southern Hemisphere populations, 31 % of the significant comparisons being between populations in the Southern Hemisphere, and 0 % coming from comparisons between populations in the North Atlantic. When the North Atlantic and Southern Hemisphere populations were grouped into two groups, the overall Fst across all loci was 0.14. For the multi-locus analysis of population structure, inspection of the average log probability of the data (ln [P(X|K)]) and the *ad hoc* ΔK statistic of Evanno et al. (2005) indicated K = 2 was the most likely number of clusters in the data (SI Fig. 1a and b). A plot of the average ln [P(X|K)] of each model as a function of K showed the likelihood increased substantially with an increase in K from 1 to 2, but increased to a lesser extent or decreased thereafter (SI Fig. 1a). Similarly, ΔK was substantially greater for a K of 2 than for any other value of K (SI Fig. 1b). The results were the same without using the location prior (results not shown).

Summary plots of Q, the estimated membership fraction for each individual, for K = 2 indicated that most individuals from the North Atlantic were assigned to cluster 1, while most individuals from the Southern Hemisphere where assigned to cluster 2 (Fig. 2). When the data were run with the location prior for all five populations, the only emergent multi-locus signal of population structure was between the North Atlantic and Southern Hemisphere. Without the location prior, the main signal was also between the North Atlantic and Southern Hemisphere, but it is clear that the two groups are connected by some migration.

Estimating genetic diversity (θ)

Estimating θ in LAMARC using all 9 loci and allowing for migration between the North Atlantic and Southern Hemisphere resulted in a posterior mean θ for the North Atlantic of 0.00096 (95 % CI 0.00048–0.0017; Table 2). From locus to locus, θ ranged from 0.0003 to 0.0026 for the North Atlantic (Table 2) and from 0.0004 to 0.0031 for the Southern Hemisphere (SI Table 2), presumably reflecting variation among loci in mutation rate or coalescent history. A comparison between the two-population migration model (North Atlantic and Southern Hemisphere) and the three-population migration model (North Atlantic, Southern Hemisphere, and North Pacific) at the 6 loci for which we had sequence data confirmed that θ in the North Atlantic was not significantly influenced by ancient migration with the North Pacific (SI Fig. 2). While the estimates of θ from the two-population, 6 locus model (MPE 0.000843, 95 % CI 0.000519–0.003239) were slightly higher then the estimates from the three-population, 6 locus model (MPE 0.000747, 95 % CI 0.000495–0.004318) (SI Fig. 2), it

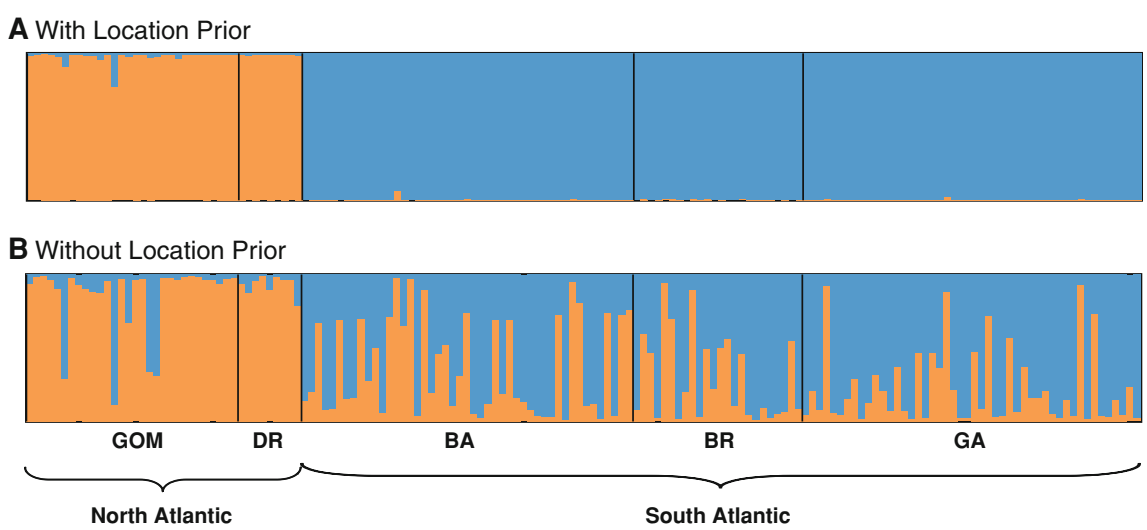

Fig. 2 Results of the multi-locus population structure analysis conducted using STRUCTURE. **a** Despite using location information for all five populations, the only emergent multi-locus signal of population structure is between the North Atlantic and the Southern Hemisphere populations of humpback whales. **b** Without the use of a location prior there is a weak, but consistent multi-locus signal of population structure between the North Atlantic and the Southern Hemisphere humpback whales

Table 2 Theta values for the North Atlantic estimated using a two population (NA and SH) migration matrix

Marker	θ	Min	Max
ACT	0.0012	0.00024	0.00353
CAT	0.0004	0.00002	0.00223
ESD	0.0011	0.00022	0.00433
FGG	0.0003	0.00001	0.00131
GBA	0.0005	0.00000	0.00323
LAC	0.0005	0.00003	0.00235
PLP*	0.0003	0.00002	0.00150
PTH	0.0018	0.00009	0.00745
RHO	0.0026	0.00019	0.01296
Posterior mean	0.0007	0.0005	0.0043

* Located on the X chromosome

is clear that inter-locus variation is much greater than variation between the two models. Thus, we conclude that the 2-population, 9 locus model adequately captured variation in θ within the NA.

Estimate of census population size from θ

Using a mutation rate of 4.40×10^{-10} (95 % CI 3.66×10^{-10}–5.29×10^{-10}) and a range of generation lengths from 12 to 24 years we calculated N_e for the North Atlantic humpback whale to be 31,900 (95 % CI 13,200–66,100). To convert N_e to N_c we estimated juvenile abundance and variation in reproductive success. We estimated juvenile abundance or the ratio of total population size to total adults to be between 1.6 and 2.0 based upon survey and catch data for humpbacks (Chittleborough 1965; Roman and Palumbi 2003). Using the ratio of N_{mature}/N_e of 2 (Nunney and Elam 1994), we multiplied the product of the two ratios by our estimate of effective population size for an estimate of census population size of 112,000 individuals (Fig. 3). Bootstrap re-sampling across the variation in mutation rate, generation lengths, the ratio of total population size to total adults and from the posterior distribution of effective size yields a 95 % CI for census size from 45,000 to 235,000.

Discussion

To improve estimates of long-term population size in the North Atlantic humpback whale, we have addressed recommendations for larger numbers of genetic loci, a better perspective on the impact of population structure, greater confidence in the mutation rate, and a greater focus on the historical timeframe of genetic population estimates (Clapham et al. 2005). Our new estimate of long-term population size of ~112,000 individuals (95 % CI 45,000–235,000) is less than half of the previous mtDNA-based estimate of ~240,000 (95 % CI 156,000–401,000) (Roman and Palumbi 2003), but is very similar to a revised population number of 150,000 (95 % CI 45,000–180,000) based on a more accurate estimate of the mutation rate (Alter and Palumbi 2009). However, the median of our most recent estimates remains far higher than the highest pre-whaling abundance estimate based upon catch data (notional upper limit: 40,000–47,000) (Smith and Pike 2009) and the discrepancy between the estimates warrants further discussion.

Population structure

Because genetic diversity within populations is strongly influenced by migration between populations, estimates of

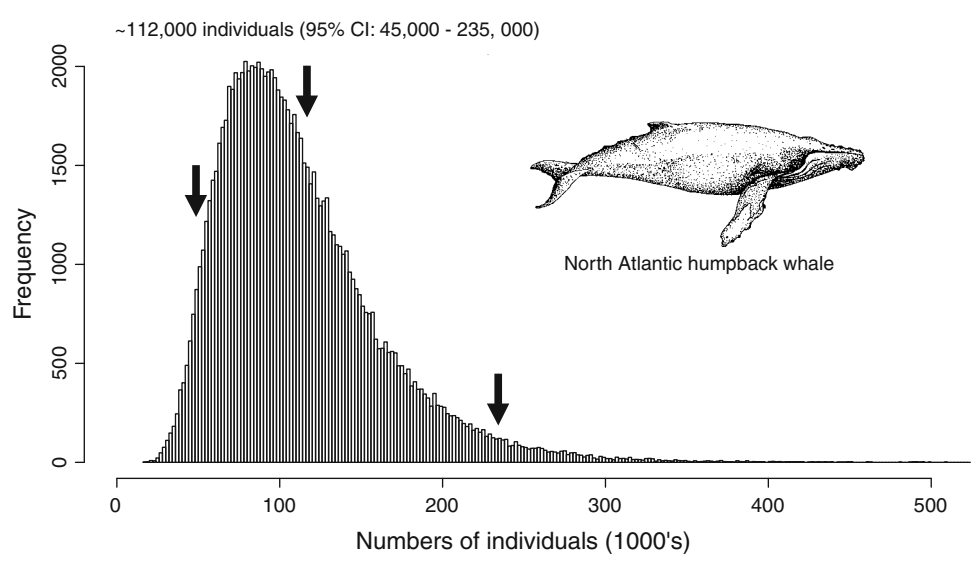

Fig. 3 Distribution of long-term census population size estimates, taking account of uncertainty in θ, mutation rate, generation time, and the ratio of total population size to total adults. The *arrows* represent the mean value and the *upper* and *lower* 95 % confidence intervals

long-term population size must account for population structure.

We re-evaluated population structure within the humpback whale based upon previous work (Baker et al. 1993; Valsecchi et al. 1997; Olavarria et al. 2007; Rosenbaum et al. 2009) and our own multi-locus analysis. Our results confirm that θ in the North Atlantic has not been significantly influenced by migration with the North Pacific (SI Fig. 2), unlike recent reports for Bowhead whales (Alter et al. 2012). Thus, our main analysis focused upon the North Atlantic and the Southern Hemisphere. While the locus-by-locus analysis revealed some signal of sub-population structure within the Southern Hemisphere (SI Table 1), the results of our within ocean basin multi-locus analysis indicate a lack of significant population structure overall, even when a strong prior for the presence of multiple sub-populations was included (Hubisz et al. 2009) (Fig. 2a). Overall, both the analysis of the average log probability of the data and the *ad hoc* ΔK statistic indicate that $K = 2$ is the most likely number of clusters in the data (SI Fig. 1). Consistent with previous research (Valsecchi et al. 1997; Olavarria et al. 2007; Rosenbaum et al. 2009), our results suggest that humpback whales within each ocean basin consist of two distinct populations connected by some migration.

One limitation with our study was the lack of samples from the eastern North Atlantic where previous research suggests the existence of a genetically distinct sub-population (Valsecchi et al. 1997). It is possible that additional samples from this region may have increased the number of distinct clusters found within the North Atlantic. However, because even small amounts of migration will cause sub-population N_e and whole population N_e to converge (Waples 2010; Hudson 1991), the absence of samples from the eastern North Atlantic sub-population is not likely to have influenced our estimate of long-term N_e. If, contrary to previous research, there is no migration between eastern and western North Atlantic feeding groups, then including samples from the eastern North Atlantic would increase our estimate of long-term population size.

Mutation rates

Attaining accurate estimates of mutation rates is a challenge common to all studies that use genetics to infer past population process (Ho et al. 2005; Emerson 2007). The difference between the original mtDNA-based estimate of ~240,000 (Roman and Palumbi 2003), the updated mtDNA-based estimate of ~150,000 (Alter and Palumbi 2009) and our multi-locus estimate of ~112,000 individuals highlights the importance of mutation rates to estimates of long-term population size. In their revised estimate, Alter and Palumbi (2009) recalibrated the control region mutation rate used in Roman and Palumbi (2003) by implementing a cytochrome b clock. Their analysis suggested that the previous mutation rate estimate was low by about two-fold because of multiple substitutions in the quickly evolving mtDNA control region. When the recalibrated control region mutation rate is employed, the mtDNA-based estimate becomes statistically indistinguishable from the multi-locus estimate of long-term population size.

Here we estimate an average mutation rate across nine nuclear loci using a phylogenetic reconstruction of the baleen phylogeny and fossil history (Jackson et al. 2009). One advantage of our multi-locus nuclear estimate is that whale nuclear DNA has far less saturation of substitutions than the mtDNA control region, and thus is far less likely to be subject to the same rate problems. Furthermore, our multi-locus approach incorporates uncertainty that results from random variation in the coalescent history of each individual locus (Rosenberg and Nordborg 2002). To adequately reflect the uncertainty in mutation rates in our final estimate of long-term population size, we bootstrap resampled across the variation in individual locus mutation rates. Thus, the multi-locus nuclear estimate that we present here should be a more robust approximation of the long-term N_e than the preceding mtDNA-only estimates.

Generation length

Uncertainty surrounding generation lengths interacts with mutation rate to determine estimates of long-term population size. Here we use a wide estimate of generation length for humpback whales ranging from 12 to 24 years (Chittleborough 1965; Roman and Palumbi 2003; Taylor et al. 2007) in order to remain consistent with previous estimates of long-term population size (Roman and Palumbi 2003). However, generation time in whales remains uncertain. Our lower bound of 12 years taken from Chittleborough (1965) is based on female age-size estimates from baleen condition, earplug layers and ovarian cycles. While Chittleborough (1965) provides the most extensive empirical data from which to estimate generation length in humpback whales, his estimates suffer from questions about age estimation (Gabriele et al. 2009; Best 2011) and whether older animals had already been culled (both of which would decrease estimates of generation length).

Taylor et al. (2007) estimated generation length for 58 cetacean species, including humpback whales, using mathematical models based on age at first reproduction and survival. They used an annual adult survival of 96 % and a first breeding age of 6 years to estimate a current generation time of 14.5 years and a stable pre-exploitation generation time of 21.5 years. While both of these estimates fall within our wide range on generation time, the

similarity between the current generation time of 14.5 years and the Chittleborough (1965) estimate of 12 years further highlights how exploitation may skew age patterns towards younger individuals. Furthermore, the absence of empirical data makes model-based estimates such as these especially sensitive to underlying assumptions. Better estimates should come from age distributions of real, unexploited populations, but such data is not readily available.

A longer estimate of humpback whale generation length would decrease our estimate of long-term population size (because the mutation rate per generation would increase). For example, if we use the estimate of 21.5 years taken from Taylor et al. (2007) we would decrease our estimate of population size to ~90,000, bringing it closer to catch based estimates of pre-whaling abundance. If we use the full range of generation times estimated through models by Taylor et al. (2007) for Baleanopterid whales (18–31 years), our mean estimate of long term population size would be ~81,000 whales (95 % CI 34,000–163,000). These results highlight the sensitivity of genetic estimates of long-term population size to estimates of generation length. However, generation time would need to be much longer than suggested by previous estimates (in excess of 64 years) in order to bring our genetic estimate of long-term population size down as low as 30,000 whales (see also Roman and Palumbi 2003). If whale generation times were actually this long, it would have far reaching implications beyond the estimation of long-term population size.

Differences in time scales

Differences been genetic and catch-based estimates of past population size may arise from the fact that genetic estimates represent an average population size over evolutionary timescales, while catch-based estimates of past population size are calculated over more recent timescales. Long-term estimates of population size based on genetic data represent the weighted harmonic mean of population size over $4N_e$ generations (e.g., up to 4,000 generations if $N_e = 1{,}000$), but with greater weight on more recent time scales (Beerli 2009). Therefore it is possible that just prior to whaling, humpback whales were less abundant than their long-term average population size. This explanation would also need to be true of gray whales in the North Pacific (Alter and Palumbi 2009) but not minke whales in the Antarctic (Ruegg et al. 2010).

In the future, it will be important to investigate more fully how past environmental variation may have influenced long-term population size in whales and whether environmental conditions just prior to whaling would have supported a population at, above, or below the long-term average abundance. This information may be especially helpful in predicting the effect of climate change on whale populations. If, for example, whale populations were generally higher during glacial maxima and lower during glacial minimum, then as the global oceans warm and ice melts, there may be a long-term decline in whale abundance. Such long-term data would be particularly useful in assessments of current and future whale conservation status.

Comparison between catch-based and genetic-based estimates of pre-whaling abundance

There has been substantial controversy surrounding the difference between genetic and catch-based estimates of pre-whaling abundance (Lubick 2003; Holt and Mitchell 2004; Clapham et al. 2005). In order to determine whether or not inaccuracies in the catch record lead to an underestimate of the number of whales before whaling, Smith and Reeves (2010) combined previously-used sources of information with additional data from archives to fill some gaps in our understanding of North Atlantic humpback whale removals. The results of their reanalysis indicate a new overall estimate of total removals that is only 6 % higher than that used previously by the IWC Scientific Committee (30,852, SE = 655). Thus, despite a reanalysis on both sides, our multi-locus estimate of average long-term population size remains higher than the pre-whaling estimate of abundance based upon catch records.

One approach to resolving these discrepancies has been population modeling (Baker and Clapham 2004). Here, historical catches, current information about reproductive rates and modern population estimates are joined together in an analytical framework that might be able to reconcile divergent views about past populations. However, population modeling performed by Punt et al. (2006) for the North Atlantic humpback shows a poor ability to explain past population crashes and current population growth. The problem stems from the fact that North Atlantic humpback whale populations can grow so quickly [6–7 % per year (Zerbini et al. 2010), that their past populations should not have collapsed at the estimated hunting rates. For example, if North Atlantic humpback whales had an original population size of 30,000 animals, and a 6–7 % annual reproductive rate at maximum sustainable yield of 64 % of the original population size (19,200 animals), then the population as a whole should have sustained a hunt of 1,152 animals a year indefinitely. Yet, data from Smith and Reeves (2010), Fig. 1 show that there has never been a recorded catch of North Atlantic humpback whales that is this high: there were only two periods of time of a few years each when the total taken was above 400 animals per year. Even though the above estimate of sustainable yield is very crude, it demonstrates the large discrepancy between the catch record and the reproductive capacity of North Atlantic whale populations.

There are several possible explanations for these discrepancies between hunting and population growth. One alternative is that the carrying capacity of the ocean to support Atlantic humpback whales might have increased 2–3 fold during the 20th century (an assumption that has not yet been supported by data or theory), and that the maximum reproductive rate of humpback whales in the 19th century was extremely low (Punt et al. 2006). Alternatively, the models would be improved if catch rates were about twice as high as suggested by Smith and Reeves (2010). Given a higher rate of catch (about 43,000–69,000 over the course of the hunt instead of 29,000, (Punt et al. 2006, Table 1), the carrying capacity of humpback whales in the North Atlantic is estimated to be about 72,000 –117,000 (Punt et al. 2006, Table 7). In general, for the 'alternate baseline' scenario that Punt et al. (2006) favor, models that suggest higher original estimates fit the data better (e.g. the negative log-likelihood values (−lnL) are closer to zero, (Punt et al. 2006, Table 6). These factors suggest that a larger historical number of humpback whales in the North Atlantic would better fit the catch data, the mathematical models and the genetic data.

The summary of these various threads of evidence is that estimates of historical abundance of North Atlantic humpback whales from catch and genetic data are converging, but remain about 2–3 fold apart. Older casual catch-based estimates of original population size from before 1990 (10,000–20,000) have been superseded by population models allowing for enhanced catch rates (20,000–46,000, Punt et al. 2006). The single locus mtDNA only genetic estimate moved from 240,000 to 150,000, after correcting for mutation rate and the addition of multiple nuclear loci resulted in an estimate of ~112,000 (95 % CI 45,000–235,000). Further declines are possible if whale reproductive life times are vastly higher than currently supposed. Regardless of the generation length employed, the lower 95 % confidence limits on multi-locus estimates of long-term population size are now much closer to the range of the population models. Further closure of these differences may depend on population trajectories of whales during climate cycles, the development of population models that correctly reflect past population trajectories, and an enhanced view of whale generation times.

Acknowledgments We thank Barry Nickel with his help with the creation of Fig. 1. This work was supported by a grant from the Lenfest Ocean Program (#2004-001492-023).

References

Alter SE, Palumbi SR (2009) Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome B in three species of baleen whales. J Mol Evol 68:97–111

Alter SE, Rynes E, Palumbi SR (2007) DNA evidence for historic population size and past ecosystem impacts of gray whales. Proc Natl Acad Sci USA 104:15162–15167

Alter ES, Rosenbaum HC, Postma LD, Whitridge P, Gaines C, Weber D, Egan MG, Lindsay M, Amato G, Dueck L, Brownell RL, Heide-Jørgensen M-P, Laidre KL, Caccone G, Hancock BL (2012) Gene flow on ice: the role of sea ice and whaling in shaping Holarctic genetic diversity and population differentiation in bowhead whales (Balaena mysticetus). Ecol Evol 2:2895–2911

Baker CS, Clapham PJ (2004) Modeling the past and future of whales and whaling. Trends Ecol Evol 19:365–371

Baker CS, Perry A, Bannister JL, Winrich MT, Abernethy RB, Camlambodikis J, Lien J, Lambertsen RH, Urban Ramirez J, Vasquez O, Clapham PJ, Alling A, O'Brien SJ, Palumbi SR (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proc Natl Acad Sci USA 90:8239–8243

Beerli P (2009) How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? In: Population genetics for animal conservation. Cambridge University Press, The Pitt Building, Trumpington St, Cambridge, pp. 42–79

Best PB (2011) A note on the age at sexual maturity of humpback whales. J Cetacean Res Manage 71–73

Braham HW (1984) The status of endangered whales: an overview. Mar Fish Rev 46:2–6

Chittleborough RG (1965) Dynamics of two populations of the humpback whale, *Megaptera novaeangllae* (Borowski). Aust J Mar Freshw Res 16:33–128

Clapham PJ, Palsboll PJ, Pastene LA, T. S, Walloe L (2005) Annex S: estimating pre-whaling abundance. J Cetacean Res Manage 7 (Suppl.), 386–387

Emerson BC (2007) Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. Syst Biol 56

Estes JA, Demaster DP, Doak DF, Williams TM, Brownell RLJ (eds) (2006) Whales, whaling and ocean ecosystems. California University Press, Berkeley

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinf (Online) 1:47–50

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925

Gabriele CM, Lockyer C, Straley JM, Jurasz CM, Kato H (2009) Sighting history of a naturally marked humpback whale (*Megaptera novaeangliae*) suggests ear plug growth layer groups are deposited annually. Mar Mammal Sci 26:443–450

Ho S, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22:1561–1568

Holt SJ, Mitchell ED (2004) Counting whales in the North Atlantic. Science 303:39–40

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Hudson RR (1991) Gene genealogies and the coalescent process. In: Oxford Surveys in Evolutionary Biology (eds. Futuyma D, Antonovics J). Oxford University Press, New York

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

IWC (2002) Annex H: Report of theSub-committee on the comprehensive assessment of North Atlantic humpback whales. J Cetacean Res Manage 4 (Suppl):230–260

IWC (2003) Report of the subcommittee on the comprehensive assessment of humpback whales. J Cetacean Res Manage 5 (Suppl.):293–323

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR (2009) Big and Slow: phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). Mol Biol Evol 11:2427–2440

Johnson JH, Wolman AA (1984) The humpback whale, *Megaptera novaeangliae*. Mar Fish Rev 46:30–37

Katona SK, Beard JA (1990) Population size, migrations and feeding aggregations of the humpback whale (*Megaptera novaeangliea*) in the western North Atlantic Ocean. Report of the International Whaling Commission, 295–305

Klinowska M (1991) Dolphins. Porpoises and Whales of the World, THe IUCN Red Data Book IUCN

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768–770

Lubick N (2003) New count of old whales adds up to big debates. Science 301:451

Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet 15:47–56

Mackintosh NA (1965) The stocks of whales. Fishing News (Books) Ltd., London

Mitchell E, Reeves RR (1983) Catch history, abundance, and present status of Northwest Atlantic humpback whales. Rep Int Whal Commn 5:153–212

Myers RA, Worm B (2003) Rapid worldwide depletion of predatory fish communities. Nature 423:280–283

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY

Nunney L, Elam DR (1994) Estimating the effective population size of conserved populations. Conserv Biol 8:175–184

Olavarria C, Baker CS, Garrigue C, Poole M, Hauser N, Caballero S, Florez-Gonzalez L, Brasseur M, Bannister J, Capella J, Clapham P, Dodemont R, Donoghue M, Jenner C, Jenner MN, Moro D, Oremus M, Paton D, Rosenbaum H, Russell K (2007) Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. Mar Ecol Prog Ser 330:257–268

Palsboll PJ, Clapham PJ, Mattila D, Larsen F, Sears R, Siegismund HR, Sigurjonsson J, Vasquez O, Arctander P (1995) Distribution of mtDNA haplotypes in North Atlantic humpback whales—the influence of behavior on population structure. Mar Ecol Prog Ser 116:1–10

Palumbi SR (1995) Nucleic acids II: The polymerase chain reaction In: Hillis D, Moritz C, Mable BK (eds) Molecular systematics. Sinauer Associates Inc., Sunderland

Palumbi SR, Roman J (2007) The history of whales read from DNA. In: Estes JA (ed) Whales, whaling and ocean ecosystems. University of California Press, Berkeley

Pauly D, Christensen V, Dalsgaard J, Froese R, Jr. FT (1998) Fishing down Marine food webs. Science 279:860–863

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Punt AE, Friday NA, Smith TD (2006) Reconciling data on the trends and abundance of North Atlantic humpback whales within a population modelling framework. J Cetacean Res Manage 8:145–159

Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. Science 301:508–510

Rosenbaum HC, Pomilla C, Mendez M, Leslie MS, Peter BB, Findlay KP, Minton G, Ersts PJ, Collins T, Engel MH, Bonatto SL, KDPG H, Meyer M, Barendse J, Thornton M, Razafindrakoto Y, Ngouessono S, Vely M, Kiszka J (2009) Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. PLoS ONE 4:e7318

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet 3:380–390

Rozas J, Sanchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR (2010) Are Antarctic minke whales unusually abundant because of 20th century whaling? Mol Ecol 19:281–291

Saiki RK, Gelfan DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Scott MJ, Goble DD, Wiens JA, Wilcove DS, Bean M, Male T (2005) Recovery of imperiled species under the Endangered Species Act: the need for a new approach. Front Ecol Environ 3:383–389

Smith TD, Pike DG (2009) The enigmatic whale: the North Atlantic humpback. NAMMCO Sci Publ 7:161–178

Smith TD, Reeves RR (2010) Historical catches of humpback whales, *Megaptera novaeangliae*, in the North Atlantic Ocean: estimates of landings and removals. Mar Fish Rev 72:1–43

Smith TD, Allen J, Clapham PJ, Hammond PS, Katona S, Larsen F, Lien J, Mattila D, Palsboll PJ, Sigurj'onsson J, Stevick PT, Oien N (1999) An ocean-basin-wide mark-recapture study of the North Atlantic humpback whale (*Megaptera novaeangliae*). Mar Mammal Sci 15:1–32

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Stevick PT, Oien N, Mattila D (1999) Migratory destinations of humpback whales from Norwegian and adjacent waters: evidence for stock identity. J Cetacean Res Manage 1:147–152

Stevick PT, Allen J, Clapham PJ, Friday N, Katona SK, Larsen F, Lien J, Mattila DK, Palsboll PJ, Sigurjonsson J, Smith TD, Oien N, Hammond PS (2003) North Atlantic humpback whale abundance and rate of increase four decades after protection from whaling. Mar Ecol Prog Ser 258:263–273

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Taylor BL, Chivers SJ, Larese J, Perrin WF (2007) Generation length and percent mature estimates for IUCN assessments of cetaceans. National Marine Fisheries Service, Southwest Fisheries Science Center

Valsecchi E, Palsboll PJ, Hale P, Glockner-Ferrari D, Ferrari M, Clapham PJ, Larsen F, Mattila D, Sears R, Sigurj'onsson J, Brown M, Cockeron P, Amos B (1997) Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). Mol Biol Evol 14:355–362

Waples RS (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Mol Ecol Resour 10:785–796

Winn HE, Reichley NE (1985) Humpback whale—*Megaptera novaeangliae* (Borowski, 1781). In: Ridgway SH, Harrison RG (eds) Handbook of marine mammals. Academic Press, London and Orlando, pp 241–273

Zerbini AN, Clapham PJ, Wade PR (2010) Assessing plausible rates of population growth in humpback whales from life-history data. Mar Biol (Berlin) 157:1225–1236