



Giants' shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis

James T Thorson¹, Jason M Cope¹, Kristin M Kleisner², Jameal F Samhour³, Andrew O Shelton³ & Eric J Ward³

¹Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA, 98112, USA; ²Sea Around Us Project, University of British Columbia, 2202 Main Mall, Vancouver, BC, Canada, V6T 1Z4; ³Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA, 98112, USA

Abstract

Meta-analysis has been an integral tool for fisheries researchers since the late 1990s. However, there remain few guidelines for the design, implementation or interpretation of meta-analyses in the field of fisheries. Here, we provide the necessary background for readers, authors and reviewers, including a brief history of the use of meta-analysis in fisheries, an overview of common model types and distinctions, and examples of different goals that can be achieved using meta-analysis. We outline the primary challenges in implementing meta-analyses, including difficulties in discriminating between alternative hypotheses that can explain the data with equal plausibility, the importance of validating results using multiple lines of evidence, the trade-off between complexity and sample size and problems associated with the use of model output. For each of these challenges, we also provide suggestions, such as the use of propensity scores for dealing with selection bias and the use of covariates to control for confounding effects. These challenges are then illustrated with examples from diverse subfields of fisheries, including (i) the analysis of the stock–recruit relationship, (ii) fisheries management, rebuilding and population viability, (iii) habitat-specific vital rates, (iv) life-history theory and (v) the evaluation of marine reserves. We conclude with our reasons for believing that meta-analysis will continue to grow in importance for these and many other research goals in fisheries science and argue that standards of practice are therefore essential.

Keywords Effect size, fisheries models, hierarchical models, meta-analysis, research synthesis

Correspondence:

James T Thorson,
Fisheries Resource
Assessment and
Monitoring Division,
Northwest Fisheries
Science Center,
National Marine
Fisheries Service,
National Oceanic and
Atmospheric Admin-
istration, 2725 Mont-
lake Blvd. East,
Seattle, WA 98112,
USA
Tel.: 971 678 5683
Fax: 206 325 6363
E-mail: James.Thor-
son@noaa.gov

Received 23 Apr
2013

Accepted 13 Sep
2013

Introduction	343
Background	343
A brief history of meta-analysis	343
Sources of variability in meta-analysis	345
Three estimation goals for meta-analysis	345
Definition	346
Challenges and Guidelines	346

Alternative hypotheses	347
Complexity vs. sample size	349
Validation using multiple lines of evidence	349
Model output and database building	349
Case-study applications	350
Example 1: recruitment	350
Example 2: rebuilding, population viability and extinction	352
Example 3: habitat-specific vital rates	353
Example 4: life-history parameters	354
Example 5: marine reserves	356
Looking forward	357
Acknowledgements	357
References	357

Introduction

The use of meta-analysis in fisheries science has steadily increased during the past two decades. In 'Standing on the shoulders of giants: learning from experience in fisheries', Hilborn and Liermann (1998) describe meta-analysis as a formal and quantitative approach to learning from past experience that can guide future research and management in fisheries science. Ram Myers popularized this approach through his many contributions to fisheries ecology and management (Myers 2001), much of which was gleaned from meta-analytic methods and a stock–recruit database that is still in use today (Ricard *et al.* 2012). These and other meta-analyses from this period revolutionized the understanding of the relationship between stock size and recruitment (Myers 2001), the likelihood of depensatory recruitment (Liermann and Hilborn 1997), and the usefulness of catch-per-unit-effort data (Harley *et al.* 2001).

Despite the important role for meta-analysis in fisheries science, there is no published list of 'best practices' for those who wish to either conduct or interpret meta-analyses. In fact, there is almost no definition of meta-analysis in fisheries science. Perhaps for this reason, the term meta-analysis is used for studies ranging from qualitative summaries of previously published results to massive statistical models that analyse data from hundreds of species.

In this study, we cast a retrospective look at what has been learned by standing on the shoulders of giants. We summarize the challenges that

continue to exist for the proper design and interpretation of meta-analytic studies, including the reliance upon model output as 'data' and the lack of globally representative databases. We then provide guidelines for the design and interpretation of future meta-analyses, and conclude by summarizing the many lessons that fisheries science has gained from meta-analytic studies since 1998. We intend to focus attention on outstanding challenges, guide future research, and aid authors, readers, and reviewers in the design and interpretation of future meta-analyses.

Background

A brief history of meta-analysis

Meta-analysis arose originally as a way to test a hypothesis using information from multiple experiments. Researchers have long sought a quantitative method to synthesize results that incorporate some measures of confidence for each study, rather than simply treating each study as a vote in favour or against a hypothesis and 'counting the votes'. An early proposal by Fisher (1925) combined the *P*-values obtained from multiple experiments. Fisher recognized that *P*-values derived from maximum likelihood statistics under the null hypothesis follow a uniform distribution, such that $-2 \cdot \ln(p)$ follows a *chi*-squared (χ^2) distribution with two degrees of freedom. Because the sum of *chi*-squared distributions is itself *chi*-squared, this forms the basis for the first meta-analytic rule for hypothesis testing with *n*-independent experiments,

$$P = -2 \sum_{i=1}^n \ln(p_i) \quad (1)$$

where P is compared with a *chi*-squared distribution with $2n$ degrees of freedom to either reject or fail to reject a null hypothesis. Fisher's method incorporates the weight of evidence for a given hypothesis from each study by means of the estimated P -value, rather than simply counting the studies that support or contradict a hypothesis. Fisher's method was followed by many other formulas for combining P -values and determining the statistical significance of the resulting statistic (Hedges and Olkin 1985).

Fisher's method is a simple example of a two-stage meta-analysis, in which an analysis is conducted for each experimental unit (i.e. study, species, etc.), and then results are synthesized in a second-stage meta-analysis model. More recent two-stage approaches include 'effect-size models', which typically extract an estimated result from each study (e.g. a response ratio for density inside and outside of an experimental treatment) and its standard errors (Gurevitch and Hedges 1999). Effect-size approaches are superior to Fisher's method and similar P -value-based approaches because they incorporate estimates of standard errors from each study as a weighting factor. This allows them to distinguish between the magnitude of an effect (an estimated value) and its precision (its standard error), which are confounded in a P -value, and hence can estimate an overall effect-size as well as test hypotheses. However, effect-size approaches have several drawbacks, including the so-called 'file-drawer problem' (i.e. published results may be biased towards statistical significance, Rosenthal 1979) and various consequences of using model output as 'data' (as discussed later).

Most recently, researchers have developed single-stage meta-analysis models as an alternative to the preceding two-stage models. These single-stage models analyse data or relatively unprocessed output from each population or study while simultaneously estimating parameters that represent the comparison among populations or studies. This single-stage approach is still susceptible to the 'file-drawer' problem, but avoids the issue of using model output as 'data'. In fisheries science, this single-stage approach to meta-analysis was popularized by Ram Myers in a series of papers beginning in the late 1990s (Myers and

Mertz 1998; Myers *et al.* 1999; Myers 2001). As one example, Harley *et al.* (2001) compare research trawl and fishery catch-per-unit-effort data to determine whether fishing efficiency changes with stock size. These and other methodological developments in the late 1990s led to the proliferation of meta-analytic research in fisheries (Fig. 1).

Rigorous meta-analyses also allow for variability among species, study sites, or whatever experimental unit is being studied. Variability that occurs randomly among experimental units is plausible in many situations, for example, different field sites may have small but repeatable differences in species densities (Osenberg *et al.* 1999) in addition to variability caused by sampling. Such variation among experimental units is ubiquitous in ecological studies due to variability in the biological processes and rates that govern ecological dynamics. Biological variability can be incorporated using 'hierarchical' models (Clark 2003; Royle and Dorazio 2008), wherein biological differences between experimental units are treated as random effects, with a prespecified parametric form for the variation. Hierarchical models have been developed for two-stage meta-analysis (Hedges and Olkin 1985) and have recently been advocated as a framework for ecological models in general (Royle and Dorazio 2008). The power of meta-analytic approaches derives largely from the

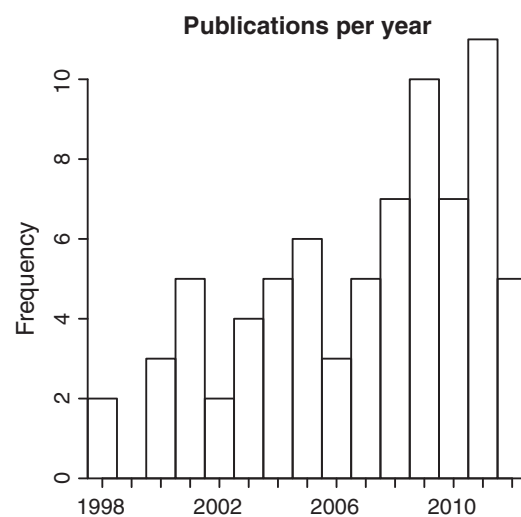


Figure 1 Number of publications per year listed in Web of Science using search terms 'fisheries' and 'meta-analysis' in their science database (queried on September 5, 2012).

flexibility of hierarchical structures to approximate nested biological systems (e.g. individual demographics within a population) and the model's ability to partition variability into meaningful units. We next discuss a general framework for considering variability in meta-analyses.

Sources of variability in meta-analysis

Meta-analytic studies seek to combine information about a process or hypothesis from multiple experimental units (i.e. studies, populations, sites, etc.). Ecological processes are often expressed using biological models, and hypotheses are evaluated by determining the ability of alternative biological hypotheses to explain available data (Hilborn and Mangel 1997).

Results may differ among experimental units due to three fundamental sources of variability (modified from Osenberg *et al.* 1999): experimental, parametric and functional variability. Experimental variability refers to random differences in study results due to sampling, as addressed by standard sampling theory and most statistical methods (Clark 2007). Parametric variability arises when different experimental units (e.g. populations) can be described by a common biological model, but the parameters of this model differ among units. For example, juvenile carrying capacity may vary among habitats for a given fish population but be well approximated by a single model for juvenile survival (Myers *et al.* 2001). Finally, functional variability refers to differences in the underlying biological processes, such that different experimental units have fundamentally different dynamics and require different biological models to be suitably approximated and interpreted. Functional variability may arise, for example, if some habitats for a fish population have population dynamics that are driven by environmental factors, while other habitats have density-dependent population dynamics (Hixon *et al.* 2012). Meta-analysis always seeks to account for the experimental variability of each study, while hierarchical meta-analysis also seeks to account for parametric variability by assuming that variation in parameters can itself be described by a distribution of possible values, that is, using random effects. However, few fisheries meta-analysis studies have addressed functional variability, and this is an important topic for future research.

Three estimation goals for meta-analysis

Meta-analyses are used for a variety of reasons, as indicated by the diversity of meta-analysis models. We will illustrate a few of the possible applications using terminology from hierarchical models (Gelman and Hill 2007; Royle and Dorazio 2008) and a simple thought experiment involving several species and a meta-analysis on a single parameter. In hierarchical models, researchers hypothesize that all observation units (e.g. populations) are influenced by a shared ecological process, but that parameters representing this process may differ among units. The parameter that varies among units is treated as a 'random effect' (i.e. a random variable that is not directly observed), and a statistical distribution of probable values (a 'hyperdistribution') is specified for the parameter. Then, parameters representing the distribution of possible values for the random effect are estimated:

$$L(\theta, \phi | \text{Data}) = \prod_{i=1}^{n_{\text{studies}}} \int_{\varepsilon_i} L(\theta_i, \varepsilon_i | \text{Data}_i) p(\varepsilon_i | \phi) \cdot d\varepsilon_i \quad (2)$$

where $L(\theta, \phi | \text{Data})$ is the likelihood of all fixed effects θ and hyperparameters ϕ given the data Data , $L(\theta_i, \varepsilon_i | \text{Data}_i)$ is the likelihood of single-study parameters θ_i and random effects ε_i for species i given data Data_i for that species, and $p(\varepsilon_i | \phi)$ is the hyperdistribution for random effects given hyperparameters. Note that the product across the studies implicitly assumes that the observed studies are independent observations of a single shared process defined by the parameters. Using this terminology, we distinguish three types of inference for meta-analyses (expanding upon those in Minte-Vera *et al.* 2005):

1. group-level inference: Meta-analyses are frequently conducted to estimate a parameter or the statistical significance of a hypothesis test for a group of experimental units in general. This *group-level* inference corresponds to the hyperparameters in a hierarchical model, that is, the mean and variance of a Gaussian hyperdistribution. The group-level mean is interpreted as an aggregate estimate of the process in general, and the standard error for the mean can be used to calculate statistical significance (i.e. whether the mean differs from a particular value in a statistically significant manner). Group-level inference also corresponds to the

- Fisher chi-squared meta-analytic test and conventional effect-size methods;
- individual-level inference: Meta-analyses can also be conducted to improve estimates for each experimental unit included in the analysis. This corresponds to improved estimates of random effects (ε_i in Equation 2) for each study and has been called the ‘Robin Hood approach’ because it borrows information from the data rich to give to the data poor (Punt *et al.* 2011). For example, single-species estimates of recruitment compensation are likely to be imprecise, and their precision will generally be improved when included in a meta-analysis (Dorn 2002) because each single-species estimate will be shrunk towards the group mean (Gelman and Hill 2007). Increased precision comes with a price, however. The shrinkage of parameter estimates towards the group-level mean implies that the parameter estimates for each population are being pulled away from the observed data for each individual population. This shrinkage is appropriate if all hierarchical parameters are ‘exchangeable’ (i.e. there is no additional information about their order or value for each individual, Gelman *et al.* 2003), but will introduce bias into individual population estimates if populations differ in unmodelled ways. This fact highlights how using hierarchical models for individual-level inference is a trade-off between increased precision and bias in the individual population estimates, and how it is important to identify ‘exchangeable’ experimental units when using a hierarchical model (see Gelman *et al.* 2003 for more details);
 - predictive inference: Hilborn and Liermann (1998) argue that the main benefit of meta-analysis is to formalize past experience for a experimental unit that is not included in any given meta-analysis. This prediction depends upon the fixed-effect hyperparameters ϕ in the hyperdistribution $p(\varepsilon_i | \phi)$, which includes the estimate of between-unit variability and also the precision of all parameter estimates (‘estimation error’).

These types of meta-analysis are summarized in Fig. 2 (derived from Minto-Vera *et al.* 2005).

Additionally, meta-analyses can be conducted using Bayesian or frequentist statistical methods. Bayesian methods are increasingly common in fisheries meta-analysis (Su *et al.* 2004; Helser *et al.* 2007; Keith *et al.* 2012; Zhou *et al.* 2012), partly

because Bayesian computational methods (i.e. Markov chain Monte Carlo) and software (e.g. BUGS and JAGS; Lunn *et al.* 2000; Plummer 2003) ease estimation of hierarchical models (see Royle and Dorazio 2008). Additionally, Bayesian approaches can incorporate prior information from previous research (Hilborn and Liermann 1998). However, Monte Carlo kernel likelihoods can be used to estimate nonlinear hierarchical models in a maximum likelihood setting (De Valpine and Hilborn 2005), and maximum likelihood approaches may be preferred when appropriate or ‘uninformative’ priors cannot be identified. Given these subtle differences, we follow De Valpine (2009) in recommending that analysts first choose an appropriate hierarchical model for meta-analysis and then decide between Bayesian and maximum likelihood implementations as a secondary consideration. Regardless of this choice, we also advocate using model diagnostic tools to assess goodness of fit (Rykiel 1996; e.g. see posterior predictive checks in Keith *et al.* 2012) and model selection to discriminate among competing hypotheses (Johnson and Omland 2004; Ward 2008), and combining goodness of fit and model selection results in the generation of new hypotheses and models (Gelman and Shalizi 2012).

Definition

Based on the preceding background, we can identify the following characteristics for meta-analyses in fisheries: (i) they summarize multiple (>5) studies or data from multiple populations and/or species that are treated as replicates for a common ecological process; (ii) they explicitly use a statistical model (i.e. they do not rely on simple vote-counting or qualitative review papers); (iii) they incorporate information or assumptions about confidence and/or precision for each study, either implicitly or explicitly. We also distinguish between hierarchical or non-hierarchical meta-analysis, where hierarchical models account for both experimental and parametric variabilities. Finally, we note the difference between single-stage and two-stage analytic approaches (as explained previously).

Challenges and guidelines

We next outline a few common challenges that are encountered in meta-analytic studies. Challenges

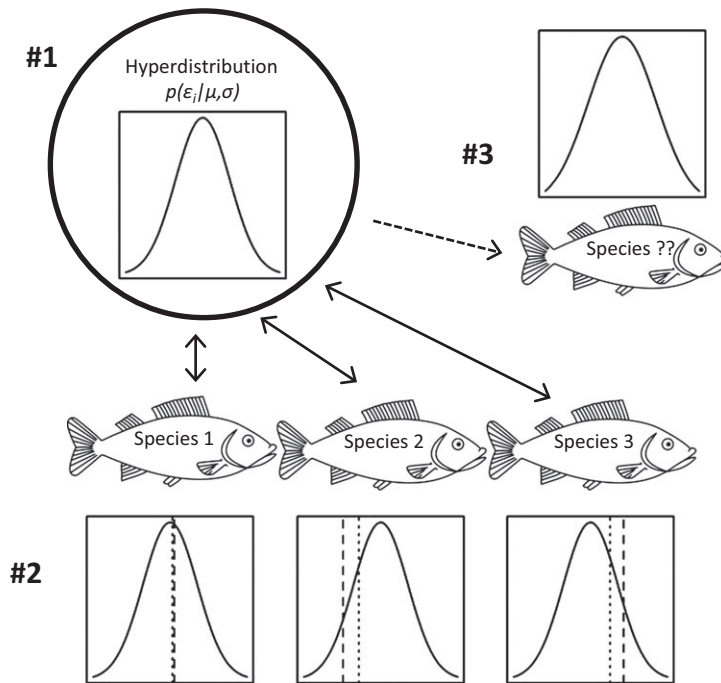


Figure 2 Schematic diagram showing a hierarchical model and the different estimation foci for various types of meta-analysis. In this schematic, there exists a hypothetical process that is shared among all four species. #1 shows an estimate of the process in general ('group-level inference') that is obtained using data for species 1–3. #2 shows the estimate for each of species 1–3 ('individual-level inference'), where the dashed line in each plot shows an estimate for each species individually, while the dotted line shows the estimate for a hierarchical model that shrinks each estimate towards a common mean. #3 shows a predictive distribution for a species with no data ('predictive inference'), which is in this case similar to the group-level distribution but also incorporates estimation uncertainty.

include the possibility of alternative hypotheses to equally explain the available data, the computational trade-off between model complexity and sample size, the importance of multiple lines of evidence in drawing strong conclusions and the drawbacks of analysing model output. This section is designed to alert researchers, readers and reviewers to common difficulties that may not be widely known or acknowledged. Where possible, we suggest general solutions to each problem. We additionally summarize these challenges and suggestions by providing six principles of fisheries meta-analysis, as summarized in Table 1.

Alternative hypotheses

Meta-analyses frequently use data that were previously collected for a different purpose ('opportunistic data'). For this reason, meta-analyses almost always have to contend with multiple hypotheses that could explain observed data. We outline two examples that highlight this issue and propose solutions for each.

1. Accounting for selection bias using propensity scores

Many large-scale meta-analyses using opportunistic data will divide experimental units between treatment and control groups, and will analyse differences between these two groups. However, it is rare that observational studies can randomly assign experimental units to treatment and control groups. In fact, experimental units in the treatment group may have been systematically selected for the treatment according to the pre-existing qualities. Therefore, any observed differences between control and treatment group could be explained either by the effect of the treatment or alternatively by the selection process for allocating units to the treatment. As one example, Melnychuk *et al.* (2012) analysed the effect of catch shares on the ability of fisheries managers to achieve management targets for 345 fish stocks. Catch shares in this case were used as a treatment, while non- or partial-catch share fisheries were used as a control. Catch shares were not randomly assigned, so the probability of having a

Table 1 Common features of meta-analyses in the fisheries case-studies reviewed in the text. Features include the complexity of the model used to conduct the meta-analysis, the sample size of individual studies on which the meta-analysis was based, the degree to which multiple lines of evidence were used to test for effects of treatments, the extent to which alternative hypotheses were considered in the meta-analysis, the number of stages used to conduct the meta-analysis (1: the study uses a hierarchical model and analyses observation-level data; 2: the study summarizes model output from prior analyses), and whether there was any accounting for experimental, parametric and functional variability in the meta-analysis design.

	Complexity	Sample size	Multiple lines of evidence	Considers alternative hypotheses	Number of stages	Experimental	Variability Parametric	Functional
Case-study 1: recruitment								
Myers <i>et al.</i> (1999)	Low	100+	No	No	2	Yes	Yes	No
Su <i>et al.</i> (2004)	High	21–100	Yes	Yes	2	Yes	Yes	Yes
Forrest <i>et al.</i> (2010)	High	5–20	No	Yes	2	Yes	Yes	Yes
Shertzer and Conn (2012)	Med	21–100	No	No	2	No	No	No
Case-study 2: rebuilding, population viability and extinction								
Dulvy <i>et al.</i> (2003)	Low	100+	No	No	2	No	No	No
Paddack <i>et al.</i> (2009)	Low	100+	No	Yes	2	Yes	No	No
Kroeker <i>et al.</i> (2010)	Medium	100+	Yes	Yes	2	Yes	Yes	No
Case-study 3: habitat-specific vital rates								
Heck <i>et al.</i> (2003)	Low	21–100	Yes	Yes	2	Yes	No	No
Minello <i>et al.</i> (2003)	Low	21–100	Yes	Yes	2	Yes	No	No
Sheridan and Hays (2003)	Low	5–20	Yes	No	2	Yes	No	No
Case-study 4: life history								
Froese and Binohlan (2000)	Low	100+	No	No	2	No	No	No
Charnov <i>et al.</i> (2013)	Low	100+	No	No	2	No	No	No
Helser <i>et al.</i> (2007)	High	21–100	No	Yes	1	Yes	Yes	No
Case-study 5: marine reserves								
Halpern (2003)	Low	21–100	Yes	No	2	No	No	No
Edgar <i>et al.</i> (2004)	Med	21–100	Yes	Yes	2	Yes	Yes	No
Babcock <i>et al.</i> (2010)	Med	5–20	Yes	Yes	2	No	No	No

catch share was confounded with the effect of catch shares. This is an example of selection bias.

To deal with the possibility of selection bias, Melnychuk *et al.* (2012) estimated the probability that each fishery would be managed using catch shares. This probability represents the ‘propensity’ for each fishery to be in the treatment group and is hence called a propensity score. Melnychuk *et al.* then used propensity score matching to compare treatment and control stocks that had an approximately equal propensity score, that is, to only compare management outcomes for stocks that had an approximately equal probability of having catch share management. Melnychuk *et al.* (2012) also used propensity score regression, where the estimated probability of having a catch share is used as a predictor in subsequent statistical comparisons and hence controls for the confounding effect of the selection process for catch shares. We recommend that propensity scores be used in similar circumstances for meta-analyses.

2. Controlling for confounding effects using additional covariates

Meta-analyses involving both experimental and observational studies will frequently have uncontrolled differences among statistical groups, in addition to any effect of statistical treatments. Uncontrolled differences between statistical groups can again be confounded with any group-level differences. As one example, Worm and Myers (2003) correlated shrimp and cod biomass to evaluate the strength of evidence for predator control of shrimp populations. However, they were concerned that apparent correlations in shrimp and cod abundance may be an artefact of responding to synchronous changes in temperature. They therefore ran several models, some of which included temperature as a covariate, and determined that the correlation remained significant even after controlling for temperature. In this way, they ruled out temperature in explaining the significant correlation between shrimp and cod

abundance. In this and other cases where a covariate of interest (i.e. treatment vs. control groups) is correlated with other confounding variables (i.e. uncontrolled differences in experimental design), we recommend that analysts include the confounding variable in the analysis to explicitly control for its effect.

Complexity vs. sample size

In addition to estimating a parameter (or parameters) representing a hypothesis of interest, meta-analytic studies frequently require the estimation of many additional parameters for each experimental unit. This leads to a computational trade-off between meta-analyses with many experimental units (e.g. species) but little complexity for each one or fewer observational units with greater complexity for each. As an example, Thorson *et al.* (2012) estimated the shape of the surplus production curve for 147 fish stocks. To accommodate data from 147 species, they used a simple four-parameter model for each species. By contrast, Thorson *et al.* (2014) analysed observation-level fishery data for 10 groundfishes to estimate correlations in recruitment off the U.S. West Coast. Although the latter analysis had fewer species and consequently fewer degrees of freedom to estimate the relationship of interest, it was able to model each individual species using the level of detail that is generally used in stock assessment models. Given this computational trade-off, we recommend that researchers make a decision about an appropriate level of detail for each experimental unit and analyse as many units as possible given that level of complexity. We also recommend that researchers seek to replicate results from global meta-analyses using regional data and greater biological detail, and vice-versa.

Validation using multiple lines of evidence

Many meta-analyses aspire to estimate a process or test a hypothesis using all available information. However, meta-analytic methods usually require that all constituent studies use a similar or comparable method and thus in practice only use a single type of information. We therefore believe that there are few meta-analyses that have been sufficiently replicated using data at different spatial/temporal scales, sampling methods or data types.

For example, studies regarding correlations in life-history parameters have generally used life-history estimates from databases such as FishBase, whose life-history parameters each have low precision but in aggregate may be highly informative. However, fewer life-history studies have attempted to estimate life-history correlations using regional or taxa-specific information, and/or higher quality information derived from fitting directly to observation-level data. These latter studies could corroborate results from global analyses and would not necessarily yield identical results, for example, if region- or taxa-specific differences are large (e.g. Thorson *et al.* 2014).

As an example of replication at different spatial and temporal scales, stock-recruit analyses typically use stock assessment estimates of stock biomass and recruitment, which are typically available with sufficient precision for analysis for <40 years. Studies conducted over a longer time horizon (e.g. >100 years) have shown much greater variability in stock biomass (reviewed by Finney *et al.* 2010), which are likely driven in part by recruitment variability at large temporal scales.

Finally, few meta-analyses have been sufficiently corroborated using multiple sampling methods. For example, the impact of marine protected areas on fish biomass is often studied using comparisons of fish densities and changes in abundance over time inside and outside reserves. However, results should ideally be corroborated using additional information, that is, large-scale estimates of biomass trajectories from stock assessment modelling and/or fishery catch-per-unit-effort data (e.g. Hamilton *et al.* 2010, 2011).

Model output and database building

Meta-analytic studies frequently use model output (i.e. parameter estimates from a preceding model) as 'data' in a subsequent model. This practice is statistically justified in some cases, for example, for Fisher's *P*-value approach. More frequently, however, the use of model output results in minor or serious violation of statistical assumptions for meta-analytic models. Maunder and Punt (2013) outline five problems with the use of model output: loss of information when converting data to model output; inconsistent assumptions between the meta-analysis model and the models that are used to generate model output; difficulties in identifying a statistical

likelihood for model output when treated as data; difficulties in representing precision for model output (i.e. due to covariance or non-normal distribution of model estimates); and reduced ability to diagnose goodness of fit for the meta-analysis model.

As a concrete example, stock assessment model estimates of recruitment and biomass have frequently been analysed to learn about the stock–recruit relationship (Liermann and Hilborn 1997; Keith *et al.* 2012). In this case, recruitment estimates are frequently treated as independent and having a common variance. Both of these assumptions are likely violated, for example, recruitment estimates for a given population are correlated among years and are likely to be more or less precisely estimated in different years. Additionally, deviations from mean estimated recruitment may not follow a lognormal distribution, so assuming that this distribution for errors may not be appropriate. We therefore propose that meta-analysis studies should, where possible, use a single model to conduct all analyses, rather than using model output as input to a meta-analysis. We acknowledge that single-stage models are not technically feasible for analyses with hundreds of species or in situations where primary data are unavailable to investigators, but believe that it is a worthwhile goal of meta-analytic research in fisheries.

Case-study applications

We next present five examples of how meta-analysis is currently used in fisheries and aquatic sciences, including (i) recruitment studies, (ii) analysis of rebuilding, population viability and extinction, (iii) analysis of habitat-specific vital rates, (iv) estimating life-history parameters and correlations, and (v) evaluating the effect of marine reserves. Each section is structured to quickly summarize the main lessons that have been learned from meta-analysis, while also offering a critique of the methods that have previously been used. Our survey of meta-analytic approaches is meant to be illustrative, not exhaustive; many excellent studies in each topic area are not discussed for brevity. We also summarize how important studies in each section have responded to the principles that we propose (Table 2). We then suggest a few improvements that could lead to progress in each field.

Example 1: recruitment

The past 15 years has seen an explosion of comparative studies of fish recruitment. The motivation for recruitment meta-analyses can be divided into two categories: (i) to provide biological information that can be used in the management of data poor stocks (Liermann and Hilborn 1997) and (ii) to synthesize information across populations and species for identifying general biological or environmental processes that determine recruitment (Myers *et al.* 1999). While the first motivation was emphasized in early meta-analyses and used as the main motivating example in Hilborn and Liermann (1998), the emphasis in recent years has shifted towards synthetic studies of recruitment processes.

The most significant contribution of recruitment meta-analyses was also one of the first. Myers *et al.* (1999) showed that despite substantial among-species variation in life-history characteristics, maximum reproductive rate was estimated to fall in a very narrow range across virtually all species examined. Subsequent meta-analyses have addressed how depensation (Liermann and Hilborn 1997; Keith *et al.* 2012), age-structure (Brunel and Boucher 2007; Venturelli *et al.* 2009), oceanography (Su *et al.* 2004; Mueter *et al.* 2007; Mantzouni *et al.* 2010) and species interactions (Minto and Worm 2012) can affect recruitment. Virtually all of these analyses use published time series of stock abundance and recruitment derived from stock assessments as input, and so treat model output as raw data.

Due to different life-history traits and fisheries data collection, Myers *et al.* (1999) decided to convert the magnitude of recruitment compensation to a dimensionless unit to enable direct comparison among stocks. Almost all subsequent analyses have used some form of standardization in recruitment and justified the standardization on the grounds that, to be comparable, the replicate units need to be expressed in identical (preferably dimensionless) units. The compensation ratio (CR; Goodyear 1977; Goodwin *et al.* 2006), spawner-per-recruit at unfished equilibrium ($SPR_F = 0$; Myers *et al.* 1999; Forrest *et al.* 2010; Mantzouni *et al.* 2010) or steepness (h ; Mace and Doonan 1988; Dorn 2002) have been used to transform recruitment compensation to a common scale. While standardization does transform populations with distinct biology to a shared scale, the process

Table 2 Six principles for conducting fisheries meta-analysis, including a brief description of the problem that they address and a recommendation for how they can be achieved.

Principle 1. Choose appropriate model complexity and sample size
Problem: Meta-analyses will frequently involve the analysis of data from tens or hundreds of experimental units. It is computationally feasible to either have complicated and biologically realistic models for fewer species or simple models for many species
Recommendation: Choose a level of model complexity that is appropriate for a given question and then select a sample size that is computationally feasible given this model choice.
Principle 2. Use multiple lines of evidence to support a hypothesis or interpretation
Problem: Meta-analysis will generally incorporate a single type of data and response metrics. However, conclusions may differ when using different response metrics
Recommendation: Identify corroborating evidence for your conclusion that uses alternative data, metrics and/or theoretical assumptions.
Principle 3. Consider alternative hypotheses
Problem: Selection bias and uncontrolled (non-randomized) variables can lead to inappropriate inference about the importance of putative factors causing a given process
Recommendation: Identify alternative hypotheses that could also explain the available data and use biological knowledge or auxiliary data to discriminate among hypotheses.
Principle 4. Strive for single-stage meta-analysis
Problem: Using an initial model to analyse data for each experimental unit and then a second model to synthesize results across species will typically violate many statistical assumptions and/or complicate interpretation of final results
Recommendation: Strive to develop a hierarchical model that uses minimally processed (observation level) data for each unit and simultaneously conducts all between-unit comparisons.
Principle 5. Account for experimental, parametric and functional variability
Problem: Ecological studies typically feature three types of variability: experimental (e.g. sampling errors), parametric (e.g. differences in demographic rates among populations) and functional (e.g. differences in compensatory mechanisms among populations)
Recommendation: Seek to incorporate all plausible types of variability, thereby ensuring that any statistical comparisons are sound.
Principle 6. Identify the desired type(s) of inference and proceed accordingly
Problem: Meta-analyses will typically focus on one of three types of inference: group level (e.g. what is the overall process), individual level (e.g. what is an updated estimate for each population) or predictive (what would an unobserved population look like). Different types of inference require different processing of model results
Recommendation: Be explicit about which inference is desired, and how reported results are matched to this type of inference.

of standardizing recruitment variables presents some generally unacknowledged complications. For example, $SPR_{F=0}$ is calculated by combining information about weight-at-age, maturity-at-age and natural mortality (see Myers *et al.* 1999; Goodwin *et al.* 2006). As all of these quantities are temporally variable and estimated with uncertainty (especially natural mortality), $SPR_{F=0}$ must also be viewed as an uncertain, estimated quantity. However, all analyses to our knowledge treat $SPR_{F=0}$ as known (e.g. Myers *et al.* 1999; Forrest *et al.* 2010; Mantzouni *et al.* 2010). The effect of treating $SPR_{F=0}$ as known when it is not is unclear, but it is certainly true that analyses based on a fixed $SPR_{F=0}$ will understate the uncertainty of recruitment parameters and affect the hierarchical model estimates. Furthermore, transforming information to a single standardized metric may impose strong but unintended assumptions into analyses (Mangel *et al.* 2010) or introduce undesirable statistical properties that impede

model estimation and biological inference (Conn *et al.* 2010).

In spite of the technical complications noted above, we believe that there is great value in the use of meta-analysis to identify shared characteristics among stocks. We highlight a study that, in our opinion, attacks the difficult problem of integrating biological complexity and statistical rigour. Su *et al.* (2004) construct a suite of models to estimate how variation in environmental conditions (e.g. sea surface temperature) affects salmon recruitment. They consider models where all populations are considered to be exchangeable (e.g. independent and identically distributed) samples from a single process, as well as hierarchical spatial models, which stipulate that geographically proximate populations are more similar than distant populations. They show that a spatial model is favoured over a non-spatial model and estimate a spatial model where the effect of sea surface temperature varies smoothly across a north-south gra-

dient. This study illustrates that it is reasonable, in many cases, to view populations that are geographically close as having more in common than geographically distant populations (i.e. the first law of geography, Tobler 1970).

Despite the number of recruitment meta-analyses conducted to date, recruitment remains a ripe area for the application of meta-analyses. Due to computing and software advances, the construction of complex hierarchical models has become relatively fast and easy. However, it is increasingly clear that stock-recruitment data at a stock-wide scale do not alone have enough information to resolve the drivers of recruitment variation. This necessitates the careful construction of models that can be linked to other data sources. We see particular potential in the incorporation of reproductive biology (e.g. the relationship between fish size and spawning output) into recruitment meta-analyses (Rothschild and Fogarty 1989; Morgan *et al.* 2007; Kuparinen *et al.* 2012). Such analyses first require that researchers break aggregated processes (e.g. the number of recruits arising from a population) into several measurable components (in this case, egg production as a function of fish size, skip-spawning as a function of fish age, sexual maturity as a function of location, age and length, etc.). Next, data must be collected for each disaggregated component, preferably in a way that uses replication among sites, years or populations (Dick 2009; Thorson *et al.* 2013; Thorson *et al.* 2013). Third, models must be developed that use parameters that are comparable among replicated sites and years (e.g. spawning frequency as a function of age; Fitzhugh *et al.* 2012). The process of disaggregating stock-wide recruitment into several measurable and replicated processes remains a ripe area for future meta-analysis.

Example 2: rebuilding, population viability and extinction

Applying meta-analysis to fish stocks is important in understanding response patterns of fish populations to fishing pressure and intrinsic population processes such as density dependence and environmental stochasticity, but also has utility for evaluating population viability and recovery. Using meta-analyses to assess rare events, such as extinction, may not be very informative in marine fish populations, because of the small sample size of populations or species that have recently gone

extinct (in recent years, most extinctions have been in freshwater systems; Reynolds *et al.* 2002). Nevertheless, fisheries meta-analyses of rebuilding stocks can illuminate which species traits are correlated with productivity and susceptibility to fishing. The core messages from these analyses are that habitat loss and overfishing are the main causes of depletion (Reynolds *et al.* 2002; Dulvy *et al.* 2003) and that depletion is much more rapid than recovery even for fast growing species (Hutchings 2000). However, we here focus on a central difficulty in meta-analyses of risk and rebuilding in fisheries, that is, how to define replicate units for analysis. Below, we explore the three approaches that have previously been used: global databases; substituting space for time in the absence of long-term monitoring; and drawing upon lessons from other taxa and/or systems.

Meta-analyses investigating patterns of the collapse and recovery of marine fish populations have most often used large databases from multiple geographical regions (Myers and Barrowman 1996, Liermann and Hilborn 1997; Hutchings 2000; Ricard *et al.* 2012). These databases typically consist of time series of spawner biomass and recruitment, and are not presented as data at the observational level, but model output from agency stock assessments. Studies in these fisheries meta-analyses may be subject to both geographical and taxa-specific selection biases. Geographically, countries in North America, Europe, Australia and New Zealand are represented well, whereas countries in South America, South-East Asia and Africa are less well represented. Similarly, fish stocks that have a history of commercial exploitation and assessment (e.g. Atlantic cod) are typically more represented than less-fished stocks. This lack of random or representative sampling could be addressed in future studies using propensity scores to control for the likelihood of a given stock being assessed and could be improved by compiling abundance information for entire regions rather than just previously assessed stocks (Pauly *et al.* 2013).

Lessons from meta-analyses built on these global fisheries stock assessment databases are typically focused on large-scale associations between species traits and management outcomes, where each population or stock is assumed to be a replicate from a shared process. For example, species that are schooling, easy to catch (both targeted and as bycatch) and have limited ranges appear to be

particularly prone to depletion (Dulvy *et al.* 2003). In another example, using a global database of stock–recruit relationships, Hutchings (2000) performed a regression meta-analysis of 90 global stocks. He demonstrated that only 8% of stocks are fully recovered 5 years after experiencing a decline of >45% and only 12% of stocks after 10 years. However, 50 of the 90 stocks in this meta-analysis belonged to the Clupeidae or Gadidae families. This lack of taxonomic representation is again typical of most global analyses of risk or rebuilding.

A second approach for defining replicate units in fisheries meta-analysis uses time-series methods that substitute space for time. In the absence of long-term datasets, spatial replicates of shorter time series can be used as a ‘space for time’ to assess the effects of fisheries management. This approach has been widely used in the analysis of ecological time-series meta-analyses (Holmes and Fagan 2002; Ward *et al.* 2009), for example, to assess the potential impacts of ocean acidification or warmer ocean conditions (Paddack *et al.* 2009; Kroeker *et al.* 2010). For example, Kroeker *et al.* (2010) were able to examine changes in growth, survival and calcification rates due to increasing acidification. They estimated these effects by extracting log-response ratios (e.g. difference in log-survival rates between normal-pH control and reduced-pH treatment groups) from published studies and analysing how they varied among taxa and between larval and adult stages. Their results showed that nearly all population processes showed negative impacts of increasing acidification, but that survival and calcification were most affected. However, future research could corroborate these results using different spatial and temporal scales in the ‘space for time’ swap.

Finally, fisheries meta-analysis can progress in identifying species recovery targets by incorporating data and conclusions from terrestrial ecology. Examples include Brook and Bradshaw (2006), which showed widespread support for density-dependent population regulation although fish species had less support than other taxa. Similarly, Gregory *et al.* (2010) examined patterns of early juvenile survival across taxa and found little support for compensatory population dynamics (Allee effects), thereby supporting previous conclusions specifically for fishes (Liermann and Hilborn 1997). However, both studies used a weight-of-evidence approach that involved either counting

the number of populations that supported a particular hypothesis or assessing the relative goodness of fit for each population individually and then subsequently averaging across species. These studies could therefore be improved by moving to a single-stage, hierarchical approach.

Example 3: habitat-specific vital rates

Marine ecology and fisheries could benefit from an improved understanding of spatial and temporal variability in demographic rates, and meta-analysis is one approach to this study. For simplicity, we restrict this example to a single demographic rate, that is, juvenile survival as it relates to habitat type and quality. For species with distinct juvenile habitat, some locations may produce disproportionately more juveniles per unit area that recruit to adult populations. Such areas are known as nursery habitats (following definition in Beck *et al.* 2001), and four factors can affect their quality, that is, greater density, juvenile survival, growth and movement to adult habitats (Beck *et al.* 2001). We found only three meta-analytic studies focusing specifically on the nursery role of marine habitats. These studies focused on seagrass (Heck *et al.* 2003; Minello *et al.* 2003), salt marsh (Minello *et al.* 2003) and mangrove (Sheridan and Hays 2003) habitats, and point to several conclusions. Of these three habitat types, the best support for a nursery role existed for seagrass habitats, and there was little indication from the meta-analyses that salt marshes and mangroves serve a nursery role. However, none of the meta-analyses were able to evaluate whether habitat-specific differences in density, growth and survival of juveniles actually influenced recruitment rates to adult populations, which is necessary to determine whether juvenile habitats are playing a nursery role (Gillanders *et al.* 2003). Similar densities or growth rates among juvenile habitats are not necessarily coupled to similar rates of successful migration to adult habitats (Beck *et al.* 2001). Nonetheless, these studies were informative as the only synthetic analyses of their kind. However, we highlight three methodological issues here: selection bias, failure to use observation-level data and uncontrolled differences in experimental design.

Selection bias is a fundamental obstacle for any study attempting to evaluate the nursery role of different juvenile habitats, because it is infeasible to randomly assign habitats to treatment and con-

control groups. Out of practical necessity, the probability of a particular location being assigned to the control treatment will be confounded with the effect of that location on juvenile densities, survival, growth and movement. Common habitats are more likely to be selected as controls, as was the case for unvegetated and seagrass habitats in the Sheridan and Hays (2003) study. This selection bias influences the outcome of any comparison with the treatment group and is inherent to the analytical framework commonly used to test the nursery habitat hypothesis (Heck *et al.* 2003; Minello *et al.* 2003). In future work, propensity score techniques may prove useful in accounting for selection bias in meta-analyses of the role of nursery habitats for fish populations.

In meta-analyses of survival in juvenile habitat, researchers generally employed effect-size models (i.e. analysed the mean and variance parameters from each primary study, as defined previously). It will remain difficult to conduct single-stage analyses until observation-level data from published studies are made publicly available on a regular and consistent basis. For example, roughly half of the studies included in the meta-analysis by Minello *et al.* (2003) were originally conducted by the authors themselves, and this situation is unlikely to change until data archiving requirements are stipulated for fishery journals. The choice to perform a two-stage meta-analysis represents a lost opportunity to improve the precision of parameter estimates from the original studies using a hierarchical modelling framework. This improvement would be particularly apparent for the primary studies in which sample sizes were small and/or estimated variance was high. In addition, a hierarchical meta-analysis would allow improved estimation of effect sizes, which provide a synthetic sense of the nursery role of alternative juvenile habitats. Finally, effect sizes generated via two-stage meta-analysis of all studies could be compared, at least qualitatively, with those from the single-stage analysis of studies where raw data were available.

A hierarchical modelling framework also makes it convenient to account explicitly for variation caused by differences in experimental design. All of the nursery habitat meta-analyses we surveyed included results from laboratory and field experiments (Heck *et al.* 2003; Minello *et al.* 2003) or from observational and experimental studies (Sheridan and Hays 2003). To their credit, in each

meta-analysis, the authors made a qualitative (Sheridan and Hays 2003) or quantitative (Heck *et al.* 2003; Minello *et al.* 2003) comparison of the influence of experimental design on estimated effect sizes. An alternative approach would have relied upon a mixed effects model that included a random effect term for the influence of the type of experimental design on the fixed treatment effect (habitat type in this case). Notably such an analysis could have been conducted using a two-stage approach in the absence of observation-level data. However, a hierarchical modelling framework would allow for more precise estimation of treatment effect sizes for individual studies by updating the estimated influence of the random effect due to experimental design based on information across all studies.

Despite these issues, we note that these three studies addressed other possible concerns. In particular, they were attuned to the potential importance of covariates that differed among primary studies. The authors accounted for factors such as geographical region, season and tidal stage through comparisons of subsets of the full data sets they analysed. The authors also addressed non-independence of model output (e.g. survival rates estimated for multiple species in a single experiment). However, none of the studies addressed all of our concerns. In particular, new meta-analyses employing hierarchical models could account for experimental design differences, covariates and non-random assignment of study sites to treatment and control groups, and would likely increase precision for effect-size estimates. Similarly, attention to functional variability across primary studies in the analysis of different factors influencing the nursery role of living habitats (e.g. the putative causes of death and growth, such as predation rates and food availability) would provide additional insight in cases where confidence intervals around effect sizes were large (e.g. Sheridan and Hays 2003).

Example 4: life-history parameters

Exploring relationships among life-history traits has a long history in fisheries science. This long history underlines both how important these parameters are to population dynamics and how difficult they are to measure. Beverton and Holt (1959) offer an early example of both compiling and analysing these traits across species groups.

Investigation of the biological basis for these relationships has also led to theory regarding life-history strategies (Stearns 1976; Adams 1980; Winemiller 1989) and life-history invariants (i.e. constants of proportionality across taxa; Charnov and Berrigan 1990).

Most life-history meta-analyses after 1997 use simple linear models applied to a mix of model output and observational data. These models tend to regress life-history traits against each other (usually in log-space) to form a predictive or at least interpretive relationship. Such an approach is intuitive and simple, but can lead to underestimation of parameter uncertainty. One major assumption is that all residual errors are in the dependent variable. The use of errors-in-variables and random effects models can help overcome such an assumption. Pascual and Iribarne (1993) warn that r^2 is a misleading measure of predictive power and that predictive intervals (the 3rd goal of meta-analysis listed previously) should explicitly be used when employing parameter estimates from fixed effects models to make inference for unstudied species.

One widely cited example of a fixed-effect approach to meta-analysis is Froese and Binohlan (2000), which presents empirical relationships for age and growth parameters, maturity and longevity. Data were culled from FishBase (www.fishbase.org, Froese 1990), which are generally taken from direct measures, but model-derived outputs are not fully excluded. The authors assume log–log relationships and include estimates of uncertainty in the form of standard errors. No predictive intervals are included. These types of relationships have become important tools, in particular for species lacking direct measures, and have been shown to be very valuable when parameterizing population dynamics (Denney *et al.* 2002) and/or simulation models (Cope and Punt 2009). However, results must be interpreted cautiously. For example, Chen and Yuan (2006) present the relationship between longevity and the intrinsic rate of population increase across 18 shark species. Both regression variables are themselves derived from empirical equations using the von Bertalanffy growth function parameters. This ‘daisy chaining’ approach to the calculation of life-history parameters is not uncommon and needs to be considered when applying the results of any subsequent analyses. Specifically, results might be appropriate for constructing a strategic simulation model but less suitable for tactical management decisions.

Life-history invariants are a special case of the log–log linear model (i.e. a linear model with x - and y -axes in log-scale) where the slope equals 1 (or -1) and the intercept is interpreted as governing a life-history relationship that is conserved across taxa. Charnov *et al.* (2013) provides a recent example as applied to estimating size-specific natural mortality from von Bertalanffy parameters for marine fishes and attribute the invariant relationship to evolutionarily derived allometric assembly rules. However, Nee *et al.* (2005) show that even random numbers can generate a slope of 1 with high r^2 , thereby questioning the previously accepted process for identifying life-history invariants.

Hierarchical models remain underrepresented in the marine fish life-history meta-analyses literature compared with conventional linear models. All hierarchical life-history models have been published after Liermann and Hilborn (1997), with the majority focusing on estimation of growth parameters. For example, Helser *et al.* (2007) conducted a meta-analysis on rockfishes (*Sebastes* spp.), a species-rich genus of marine fishes. Their model included experimental variability in terms of individual length-at-age measurements, as well as parametric variation between species and for parameters within a given species. The authors also included gender, depth and size at maturity as covariates to control for these confounding effects, and estimated growth parameters for each study species as well as for a species with unknown growth parameters. They compare the results of this model with linear models fit to each species in turn. They found that estimates were comparable whether they were fit individually or in the hierarchical model, but that the hierarchical model produced lower parameter variance and correlation among parameters. The hierarchical structure was able to provide improved estimates of growth curves for data-poor species, as well as for species with no growth data.

Life-history traits remain a ripe area for application of meta-analysis in part due to the continuing development of life-history databases. However, results from life-history meta-analyses should be applied cautiously, especially when it comes to out-of-sample extrapolation. Meta-analytical approaches should not necessarily assume that the species included in the analysis appropriately represent species not included in the analysis. As one example, the Charnov *et al.* (2013) method of esti-

mating natural mortality may be appropriate to many marine fishes, but it is inappropriate for many shark species due to large differences in life-history relationships between fishes and elasmobranchs.

Example 5: marine reserves

Marine reserves are a popular management strategy for exploited species. However, many reserves are created without an understanding of whether they effectively increase fish density, biomass, size and other biological characteristics. To address this information gap, several large meta-analyses have compared multiple reserves to quantify differences in effect size and recovery rate as a function of location and longevity. Overall, meta-analyses have found that marine reserves provide positive benefits both to species and habitats. However, the reliability of these results has been questioned due to the lack of comparability between measures of effectiveness, site location, model designs and the confounding of model assumptions with demonstrated facts. In this last case-study, we discuss some of the key meta-analytic studies on marine reserve effectiveness with respect to their design and potential sources of bias.

The published studies on marine reserve effectiveness prior to 2005 tended to be control-impact designs where sites within the reserves were compared with adjacent sites outside the reserve. In particular, the studies by Halpern and Warner (2002) and Halpern (2003) included 89 independent studies of marine reserve effects, only 17 of which included 'before-after' measurements, and only nine of which included both 'before-after' and 'control-impact' measurements. One of the first studies to look critically at the differences between reserves was a meta-analysis of reserve 'zones' within the Galapagos Marine Reserve (Edgar *et al.* 2004). In this study, a comparison was made between 62 tourism sites (no fishing allowed), 45 fishing zones (no tourism allowed, 'no-take') and 9 mixed-use zones. This study found that densities of sea cucumbers, the most valuable fishery resource in the Galapagos, were three times higher in the areas open to fishing. This surprising result was attributed to selection bias (as discussed previously) in the placement of the reserve zones due to socio-political factors, for example, fishers push for larger no-take zones in resource poor areas, whereas divers push for protection of areas with

features of interest. The authors point out that this selection bias, which is not typically included in marine reserve studies, could serve to confound the results from meta-analyses that incorporate results from such 'control-impact' studies.

Additionally, before-after control-impact studies only capture the direct effects of marine reserve implementation that produce an absolute increase in abundance, size or biomass. They cannot capture indirect effects due to reserves, which may be a result of species interactions. Babcock *et al.* (2010) attribute this to the fact that trophic interactions are determined by absolute values (of abundance, size or biomass), not relative differences between reserves and fished areas. For example, ratios between reserves and fished areas may not capture reserve effects if community assemblages are protected within the reserve while targeted species decline over time in fished areas, possibly at a faster rate than prior to reserve implementation due to displaced fishing effort. Therefore, the ratio approach may actually provide fewer insights into the temporal dynamics and associated variation in reserve effects than continuous time series after closure. Hence, Babcock *et al.* (2010) used continuous time series of raw data for each reserve to explore the temporal dynamics that may produce changes in marine reserves as opposed to ratios of reserve to non-reserve values. Using raw data, they avoided issues related to model output. They found that the detection of direct effects on target species occurred over relatively short intervals after reserve implementation (5.13 ± 1.9 years), while the detection of indirect effects on non-target species occurred after significantly longer periods (13.1 ± 2.0 years). In general, target species showed initial direct effects, but there was considerable variation in species trajectories over time (parametric variation).

Willis *et al.* (2003) noted several key aspects of current research on the potential for positive effects of reserves on marine species. First, while the number of empirical studies on marine reserves increased between 1991 and 2001, this lagged behind the number of review studies. Second, the number of theoretical studies of marine reserve benefits also increased greatly since 1997. However, they noted a lack of published evidence to empirically judge these models and their assumptions. Third, the fact that there were so many new models and reviews had led many to

take model assumptions as accepted theory. They concluded that while there may be many benefits garnered from marine reserve establishment, often, 'intuitive speculations' appear as 'logically true assertions' in the literature. These lessons could also be drawn from meta-analyses regarding recruitment, life-history invariants or many other fields where theoretical arguments and meta-analyses have generated conclusions that may well be controverted when additional data are collected and analysed using appropriate hierarchical models.

Looking forward

In this study, we have provided a general background and several key distinctions among the common types of meta-analyses seen in fisheries. We listed the common and often unacknowledged problems that occur in meta-analysis, including difficulties in distinguishing between alternative hypotheses, the importance of validating results using multiple lines of evidence, and the trade-off between complexity and sampling size. We then showed how these common problems are encountered within different research threads in fisheries science. We now conclude by providing our perspective on the future of fisheries meta-analysis.

We believe that meta-analysis has an important role to play in fisheries science for three reasons. First, fisheries questions are often characterized by a large spatial and temporal scales, and this often necessitates the use of opportunistically collected data. Second, marine fishes are often difficult to observe, and therefore, surveys of fish abundance often have large measurement errors (e.g. Thorson and Ward 2013). For this reason, information might be scarce for a single site and/or population, which necessitates the combination of information from multiple experimental units. Third, fisheries questions are urgent. Research results are frequently used for management of recreational and commercial fishers, and we therefore are obligated to pursue accurate and highly precise estimates for complicated processes. In many cases, these goals can only be attained by synthesizing information from multiple species.

We also note that fisheries research, like the rest of population ecology, is being revolutionized by several important trends. These include the increasing prevalence of rich and global databases, that is, FishBase (Froese 1990) for life-history

traits, the RAM Legacy Stock Assessment Database (Ricard *et al.* 2012) for stock-recruitment and management information and the Sea Around Us project (Pauly 2007) for productivity, fishing effort, spatial catch and many other data products. Also important is the expanded feasibility and ease of building complex and detailed meta-analysis models. This is helped by improved computational power, but also by improved statistical software such as JAGS (Plummer 2003) for Bayesian modelling and AD Model Builder (Fournier *et al.* 2012) for maximum likelihood models.

However, these trends also provide greater leeway for poorly planned and implemented meta-analysis. Taking a retrospective look at meta-analysis since its popularization by Hilborn and Liermann's (1998) 'Standing on the shoulders of giants...', we see that many problems continue to occur for meta-analyses spanning the range from recruitment studies to evaluation of management actions. We therefore propose that authors, readers and reviewers take note of the many challenges and solutions that we have noted, and hope that this will improve the quality of meta-analyses going forward, with the potential to advance nearly every field in fisheries science.

Acknowledgements

The authors are grateful for many discussions with colleagues that contributed to the quality of this work, including helpful advice from J. Hastie, M. McClure, and W. Satterthwaite. K. Kleisner was supported by the *Sea Around Us* project, a collaboration between the University of British Columbia and the Pew Environment Group and Conservation International. The manuscript was improved by comments from two anonymous reviews and Tony Pitcher.

References

- Adams, P.B. (1980) Life history patterns in marine fishes and their consequences for fisheries management. *Fishery Bulletin* **78**, 1–12.
- Babcock, R.C., Shears, N.T., Alcala, A.C. *et al.* (2010) Decadal trends in marine reserves reveal differential rates of change in direct and indirect effects. *Proceedings of the National Academy of Sciences* **107**, 18256–18261.
- Beck, M.W., Heck Jr, K.L., Able, K.W. *et al.* (2001) The identification, conservation, and management of estua-

- rine and marine nurseries for fish and invertebrates. *BioScience* **51**, 633–641.
- Beverton, R. and Holt, S. (1959) A review of the life-spans and mortality rates of fish in nature, and their relation to growth and other physiological characteristics. In: *Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing)* (eds G.E.W. Wolstenholme and M. O'Connor). J. and A. Churchill Ltd., London, pp. 142–177.
- Brook, B.W. and Bradshaw, C.J. (2006) Strength of evidence for density dependence in abundance time series of 1198 species. *Ecology*, **87**, 1445–1451.
- Brunel, T. and Boucher, J. (2007) Long-term trends in fish recruitment in the north-east Atlantic related to climate change. *Fisheries Oceanography* **16**, 336–349.
- Charnov, E.L. and Berrigan, D. (1990) Dimensionless numbers and life history evolution: age of maturity versus the adult lifespan. *Evolutionary Ecology* **4**, 273–275.
- Charnov, E.L., Gislason, H. and Pope, J.G. (2013) Evolutionary assembly rules for fish life histories. *Fish and Fisheries* **14**, 213–224.
- Chen, P. and Yuan, W. (2006) Demographic analysis based on the growth parameter of sharks. *Fisheries research* **78**, 374–379.
- Clark, J.S. (2003) Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology* **84**, 1370–1381.
- Clark, J.S. (2007) *Models for Ecological Data: An Introduction*. Princeton University Press, Princeton, NJ.
- Conn, P.B., Williams, E.H. and Shertzer, K.W. (2010) When can we reliably estimate the productivity of fish stocks? *Canadian Journal of Fisheries and Aquatic Sciences* **67**, 511–523.
- Cope, J.M. and Punt, A.E. (2009) Length-based reference points for data-limited situations: applications and restrictions. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **1**, 169–186.
- De Valpine, P. (2009) Shared challenges and common ground for Bayesian and classical analysis of hierarchical statistical models. *Ecological Applications* **19**, 584–588.
- De Valpine, P. and Hilborn, R. (2005) State-space likelihoods for nonlinear fisheries time-series. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 1937–1952.
- Denney, N.H., Jennings, S. and Reynolds, J.D. (2002) Life-history correlates of maximum population growth rates in marine fishes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 2229–2237.
- Dick, E.J. (2009) *Modeling the Reproductive Potential of Rockfishes (Sebastes spp.)*. Available at: <http://users.soe.ucsc.edu/~msmangel/Dick%20Thesis.pdf> (accessed 25 September 2013).
- Dorn, M.W. (2002) Advice on West Coast rockfish harvest rates from Bayesian meta-analysis of stock-recruit relationships. *North American Journal of Fisheries Management* **22**, 280–300.
- Dulvy, N.K., Sadovy, Y. and Reynolds, J.D. (2003) Extinction vulnerability in marine populations. *Fish and Fisheries* **4**, 25–64.
- Edgar, G.J., Bustamante, R.H., Farina, J.M., Calvopina, M., Martinez, C. and Toral-Granda, M.V. (2004) Bias in evaluating the effects of marine protected areas: the importance of baseline data for the Galapagos Marine Reserve. *Environmental Conservation* **31**, 212–218.
- Finney, B.P., Alheit, J., Emeis, K.-C., Field, D.B., Gutiérrez, D. and Struck, U. (2010) Paleocological studies on variability in marine fish populations: a long-term perspective on the impacts of climatic change on marine ecosystems. *Journal of Marine Systems* **79**, 316–326.
- Fisher, S.R.A. (1925) *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- Fitzhugh, G.R., Shertzer, K.W., Kellison, G.T. and Wyanski, D.M. (2012) Review of size-and age-dependence in batch spawning: implications for stock assessment of fish species exhibiting indeterminate fecundity. *Fishery Bulletin* **110**, 413–425.
- Forrest, R.E., McAllister, M.K., Dorn, M.W., Martell, S.J.D. and Stanley, R.D. (2010) Hierarchical Bayesian estimation of recruitment parameters and reference points for Pacific rockfishes (*Sebastes* spp.) under alternative assumptions about the stock-recruit function. *Canadian Journal of Fisheries and Aquatic Sciences* **67**, 1611–1634.
- Fournier, D.A., Skaug, H.J., Ancheta, J. et al. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* **27**, 1–17.
- Froese, R. (1990) FishBase: an information system to support fisheries and aquaculture research. *ICLARM Fishbyte* **8**, 21–24.
- Froese, R. and Binohlan, C. (2000) Empirical relationships to estimate asymptotic length, length at first maturity and length at maximum yield per recruit in fishes, with a simple method to evaluate length frequency data. *Journal of Fish Biology* **56**, 758–773.
- Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, UK.
- Gelman, A. and Shalizi, C.R. (2012) Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology* **66**, 8–38.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003) *Bayesian Data Analysis*, 2nd edn. Chapman & Hall, Boca Raton, FL.
- Gillanders, B., Able, K.W., Brown, J.A., Eggleston, D.B. and Sheridan, P.F. (2003) Evidence of connectivity between juvenile and adult habitats for mobile marine fauna: an important component of nurseries. *Marine Ecology Progress Series* **247**, 281–295.

- Goodwin, N.B., Grant, A., Perry, A.L., Dulvy, N.K. and Reynolds, J.D. (2006) Life history correlates of density-dependent recruitment in marine fishes. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 494–509.
- Goodyear, C.P. (1977) Assessing the impact of power plant mortality on the compensatory reserve of fish populations.
- Gregory, S.D., Bradshaw, C.J., Brook, B.W. and Courchamp, F. (2010) Limited evidence for the demographic Allee effect from numerous species across taxa. *Ecology* **91**, 2151–2161.
- Gurevitch, J. and Hedges, L.V. (1999) Statistical issues in ecological meta-analyses. *Ecology* **80**, 1142–1149.
- Halpern, B.S. (2003) The impact of marine reserves: do reserves work and does reserve size matter? *Ecological Applications* **13**, 117–137.
- Halpern, B.S. and Warner, R.R. (2002) Marine reserves have rapid and lasting effects. *Ecology Letters* **5**, 361–366.
- Hamilton, S.L., Caselle, J.E., Malone, D.P. and Carr, M.H. (2010) Incorporating biogeography into evaluations of the Channel Islands marine reserve network. *Proceedings of the National Academy of Sciences* **107**, 18272–18277.
- Hamilton, S.L., Wilson, J.R., Ben-Horin, T. and Caselle, J.E. (2011) Utilizing spatial demographic and life history variation to optimize sustainable yield of a temperate sex-changing fish. *PLoS ONE* **6**, e24580.
- Harley, S.J., Myers, R.A. and Dunn, A. (2001) Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1760–1772.
- Heck, K.L., Hays, G. and Orth, R.J. (2003) Critical evaluation of the nursery role hypothesis for seagrass meadows. *Marine Ecology Progress Series* **253**, 123–136.
- Hedges, L.V. and Olkin, I. (1985) *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL.
- Helser, T.E., Stewart, I.J. and Lai, H.L. (2007) A Bayesian hierarchical meta-analysis of growth for the genus *Sebastes* in the eastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 470–485.
- Hilborn, R. and Liermann, M. (1998) Standing on the shoulders of giants: learning from experience in fisheries. *Reviews in Fish Biology and Fisheries* **8**, 273–283.
- Hilborn, R. and Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data*, 1st edn. Princeton University Press, Princeton, NJ.
- Hixon, M.A., Anderson, T.W., Buch, K.L., Johnson, D.W., McLeod, J.B. and Stallings, C.D. (2012) Density dependence and population regulation in marine fish: a large-scale, long-term field manipulation. *Ecological Monographs* **82**, 467–489.
- Holmes, E.E. and Fagan, W.F. (2002) Validating population viability analysis for corrupted data sets. *Ecology* **83**, 2379–2386.
- Hutchings, J.A. (2000) Collapse and recovery of marine fishes. *Nature* **406**, 882–885.
- Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**, 101–108.
- Keith, D.M., Hutchings, J.A. and Hilborn, R. (2012) Population dynamics of marine fishes at low abundance. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 1150–1163.
- Kroeker, K.J., Kordas, R.L., Crim, R.N. and Singh, G.G. (2010) Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. *Ecology Letters* **13**, 1419–1434.
- Kuparinen, A., Mäntyniemi, S., Hutchings, J.A. and Kuikka, S. (2012) Increasing biological realism of fisheries stock assessment: towards hierarchical Bayesian methods. *Environmental Reviews* **20**, 135–151.
- Liermann, M. and Hilborn, R. (1997) Depensation in fish stocks: a hierarchical Bayesian meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1976–1984.
- Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.
- Mace, P.M. and Doonan, I.J. (1988) *A Generalised Bioeconomic Simulation Model for Fish Population Dynamics*. New Zealand Ministry of Agriculture and Fisheries, Wellington.
- Mangel, M., Brodziak, J. and DiNardo, G. (2010) Reproductive ecology and scientific inference of steepness: a fundamental metric of population dynamics and strategic fisheries management. *Fish and Fisheries* **11**, 89–104.
- Mantzouni, I., Sørensen, H., O'Hara, R.B. and MacKenzie, B.R. (2010) Hierarchical modelling of temperature and habitat size effects on population dynamics of North Atlantic cod. *ICES Journal of Marine Science: Journal du Conseil* **67**, 833–855.
- Maunder, M.N. and Punt, A.E. (2013) A review of integrated analysis in fisheries stock assessment. *Fisheries Research* **142**, 61–74.
- Melnichuk, M.C., Essington, T.E., Branch, T.A. et al. (2012) Can catch share fisheries better track management targets? *Fish and Fisheries* **13**, 267–290.
- Minello, T.J., Able, K.W., Weinstein, M.P. and Hays, C.G. (2003) Salt marshes as nurseries for nekton: testing hypotheses on density, growth and survival through meta-analysis. *Marine Ecology Progress Series* **246**, 39–59.
- Minte-Vera, C.V., Branch, T.A., Stewart, I.J. and Dorn, M.W. (2005) Practical application of meta-analysis results: avoiding the double use of data. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 925–929.

- Minto, C. and Worm, B. (2012) Interactions between small pelagic fish and young cod across the North Atlantic. *Ecology* **93**, 2139–2154.
- Morgan, M.J., Shelton, P.A. and Brattey, J. (2007) Age composition of the spawning stock does not always influence recruitment. *Journal of Northwest Atlantic Fishery Science* **38**, 1–12.
- Mueter, F.J., Boldt, J.L., Megrey, B.A. and Peterman, R.M. (2007) Recruitment and survival of Northeast Pacific Ocean fish stocks: temporal trends, covariation, and regime shifts. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 911–927.
- Myers, R.A. (2001) Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES Journal of Marine Science: Journal du Conseil* **58**, 937–951.
- Myers, R.A. and Barrowman, N.J. (1996) Is fish recruitment related to spawner abundance? *Fishery Bulletin* **94**, 707–724.
- Myers, R. and Mertz, G. (1998) Reducing uncertainty in the biological basis of fisheries management by meta-analysis of data from many populations: a synthesis. *Fisheries Research* **37**, 51–60.
- Myers, R.A., Bowen, K.G. and Barrowman, N.J. (1999) Maximum reproductive rate of fish at low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 2404–2419.
- Myers, R.A., MacKenzie, B.R., Bowen, K.G. and Barrowman, N.J. (2001) What is the carrying capacity for fish in the ocean? A meta-analysis of population dynamics of North Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1464–1476.
- Nee, S., Colegrave, N., West, S.A. and Grafen, A. (2005) The illusion of invariant quantities in life histories. *Science* **309**, 1236–1239.
- Osenberg, C.W., Sarnelle, O., Cooper, S.D. and Holt, R.D. (1999) Resolving ecological questions through meta-analysis: goals, metrics, and models. *Ecology* **80**, 1105–1117.
- Paddack, M.J., Reynolds, J.D., Aguilar, C. et al. (2009) Recent region-wide declines in Caribbean reef fish abundance. *Current Biology* **19**, 590–595.
- Pascual, M.A. and Iribarne, O.O. (1993) How good are empirical predictions of natural mortality? *Fisheries Research* **16**, 17–24.
- Pauly, D. (2007) The Sea Around Us Project: documenting and communicating global fisheries impacts on marine ecosystems. *Ambio* **34**, 290–295.
- Pauly, D., Hilborn, R. and Branch, T.A. (2013) Fisheries: does catch reflect abundance? *Nature* **494**, 303–306.
- Plummer, M. (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. Vienna, Austria.
- Punt, A.E., Smith, D.C. and Smith, A.D. (2011) Among-stock comparisons for improving stock assessments of data-poor stocks: the “Robin Hood” approach. *ICES Journal of Marine Science: Journal du Conseil* **68**, 972.
- Reynolds, J.D., Dulvy, N.K. and Roberts, C.M. (2002) 15 Exploitation and other threats to fish conservation. *Handbook of Fish Biology and Fisheries* **2**, 319.
- Ricard, D., Minto, C., Jensen, O.P. and Baum, J.K. (2012) Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries* **13**, 380–398.
- Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychological bulletin* **86**, 638.
- Rothschild, B.J. and Fogarty, M.J. (1989) Spawning-stock biomass: a source of error in recruitment/stock relationships and management advice. *Journal du Conseil: ICES Journal of Marine Science* **45**, 131–135.
- Royle, J.A. and Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*, 1st edn. Academic Press, London.
- Rykiel, E.J. (1996) Testing ecological models: the meaning of validation. *Ecological modelling* **90**, 229–244.
- Sheridan, P. and Hays, C. (2003) Are mangroves nursery habitat for transient fishes and decapods? *Wetlands* **23**, 449–458.
- Shertzer, K.W. and Conn, P.B. (2012) Spawner-recruit relationships of demersal marine fishes: prior distribution of steepness. *Bulletin of Marine Science* **88**, 39–50.
- Stearns, S.C. (1976) Life-history tactics: a review of the ideas. *The Quarterly Review of Biology* **51**, 3–47.
- Su, Z., Peterman, R.M. and Haeseker, S.L. (2004) Spatial hierarchical Bayesian models for stock-recruitment analysis of pink salmon (*Oncorhynchus gorbuscha*). *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 2471–2486.
- Thorson, J.T. and Ward, E. (2013) Accounting for space-time interactions in index standardization models. *Fisheries Research* **143**, 426–433.
- Thorson, J.T., Cope, J., Branch, T. and Jensen, O. (2012) Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 1556–1568.
- Thorson, J., Taylor, I., Stewart, I.J. and Punt, A.E. (2013) Using a recruitment-linked multispecies stock assessment model to estimate common trends in recruitment for U.S. West Coast groundfishes. *Marine Ecology Progress Series* **483**, 245–256.
- Thorson, J.T., Scheuerell, M.D., Buhle, E. and Copeland, T. (2013) Spatial variation buffers temporal fluctuations in early juvenile survival for an endangered Pacific salmon. *Journal of Animal Ecology* **83**, 157–167.

- Thorson, J.T., Taylor, I., Stewart, I.J. and Punt, A.E. (2014) Rigorous meta-analysis of life history correlations by simultaneously analyzing multiple population dynamics models. *Ecological Applications* **24**, 315–326.
- Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46**, 234–240.
- Venturelli, P.A., Shuter, B.J. and Murphy, C.A. (2009) Evidence for harvest-induced maternal influences on the reproductive rates of fish populations. *Proceedings of the Royal Society B: Biological Sciences* **276**, 919–924.
- Ward, E.J. (2008) A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. *Ecological Modelling* **211**, 1–10.
- Ward, E.J., Chirakkal, H., González-Suárez, M., Aurióles-Gamboa, D., Holmes, E.E. and Gerber, L. (2009) Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. *Journal of Applied Ecology* **47**, 47–56.
- Willis, T.J., Millar, R.B., Babcock, R.C. and Tolimieri, N. (2003) Burdens of evidence and the benefits of marine reserves: putting Descartes before des horse? *Environmental conservation* **30**, 97–103.
- Winemiller, K.O. (1989) Patterns of variation in life history among South American fishes in seasonal environments. *Oecologia* **81**, 225–241.
- Worm, B. and Myers, R.A. (2003) Meta-analysis of cod-shrimp interactions reveals top-down control in oceanic food webs. *Ecology* **84**, 162–173.
- Zhou, S., Yin, S., Thorson, J., Smith, T. and Fuller, M. (2012) Linking fishing mortality reference points to life history traits: an empirical study. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 1292–1301.