New Approaches to Marine Conservation Through the Scaling Up of Ecological Data — Source link

Graham J. Edgar, Amanda E. Bates, Tomas J. Bird, Alun H. Jones ...+3 more authors

Institutions: University of Tasmania, National Oceanography Centre, Southampton, University of Southampton, University of Sheffield ...+1 more institutions

Published on: 08 Jan 2016 - Annual Review of Marine Science (Annual Reviews)

Topics: Marine conservation and Ecosystem management

Article:

Edgar, G.J., Bates, A.E., Bird, T.J. et al. (4 more authors) (2016) New Approaches to Marine Conservation Through the Scaling Up of Ecological Data. Annual Review of Marine Science, 8. pp. 435-461. ISSN 1941-1405

https://doi.org/10.1146/annurev-marine-122414-033921

New Approaches to Marine Conservation Through the Scaling Up of Ecological Data

Submission to Annual Review of Marine Science

Graham J. Edgar[1*], Amanda Bates[2], Tom Bird[3], Alun H. Jones[4], Stuart Kininmonth[5], Rick D. Stuart-Smith[1], Thomas J. Webb[4]

[1]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia
g.edgar@utas.edu.au
Rick.StuartSmith@utas.edu.au
[2]National Oceanography Centre, University of Southampton Waterfront Campus, Southampton SO14 3ZH, UK
A.E.Bates@soton.ac.uk
[3]Department of Geography and the Environment, University of Southampton, Southampton, SO17 1BJ, UK
tomasbird@gmail.com
[4]Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
AHJones1@sheffield.ac.uk
t.j.webb@sheffield.ac.uk
[5]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, SE-106 91 Stockholm
stuart@kininmonth.com.au

Running title: Global-scale ecological monitoring

[*]Corresponding author: Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia g.edgar@utas.edu.au +61 427 811 699

TABLE OF CONTENTS

Comment [GE1]: Note to editor: Ideally, these should each be boxed separately from text

Keywords

Biological diversity, citizen science, ecological monitoring, ecosystem management, macro-ecology, underwater visual census

Abstract

In an era of rapid global change, conservation managers urgently need improved tools for countering declining ecosystem condition. This need is particularly acute in the marine realm, where threats are out-of-sight, inadequately mapped, cumulative, and poorly understood, thereby generating impacts that are inefficiently managed. Recent advances in macroecology, statistics, and the compilation of global data will play a central role in improving conservation outcomes, provided that global, regional and local data streams can be integrated to produce locally-relevant and interpretable outputs. Progress will be assisted by 1) expanded rollout of systematic surveys that quantify the distribution of marine species, including through assistance of citizen scientists, 2) improved understanding of consequences of threats through application of recently-developed statistical techniques to species' distributional data and associated environmental and socio-economic covariates, 3) development of reliable ecological indicators for accurate and comprehensible tracking of threats, and 4) improved data-handling and communication tools.

INTRODUCTION

New approaches to the collection and analysis of data have immense potential to transform conservation management, including through application of novel tools borrowed from such diverse fields as genomics, statistics, socio-economics, philosophy, demography and biogeography. For marine environments, this revolution will arguably be led by the application of 'big' ecological data, given the hidden nature of aquatic realms and the paucity of existing data. Until recently, data describing the distribution and abundance of marine organisms have been sparse, disconnected, largely qualitative, and expensive to collect. Without dense and reliable information, the tools to quantify environmental health, or adequate techniques for synthesising and visualising that information, managers have had little alternative other than to act intuitively when allocating limited resources to minimise threats to marine life.

The logistical challenges associated with collecting data on marine organisms and environments has in the past resulted in a focus in ecology on small-plot manipulative experiments in intertidal and shallow subtidal depths, where the scale of the study is limited to access by car or small boats. More recently, investigations have progressed towards study of large-scale (i.e. national or global) macroecological processes, with individual surveys undertaken at a variety of local scales, in part depending on the complexity of observations (Fig. 1). These multiple scales of investigation have been key to the development of a deeper understanding of marine ecology. While local-scale investigations frequently show strong influences of specific factors, such relationships often break down when re-assessed over regional scales. By contrast, numerous examples exist of strong relationships between ecological patterns and environmental covariates that emerge clearly only when multinational-scale data are considered (e.g., Mora et al 2011, Webb et al 2009). Thus, alongside the traditional manipulative plot experiment, regional- to global-scale data are now recognised as fundamental to progress in ecology (Keith et al 2012, Kerr et al 2007). Excitingly, the costs associated with collecting and analysing such data have declined dramatically due to the development of improved tools for the collection, communication, and analysis of data, as well as the lowering of key logistical barriers such as cost of air travel, acquisition and training of volunteer workers, and open-sourcing of software and data.

Because of difficulties monitoring species across broad scales, much effort to the present has been directed at using habitat and aggregate assemblage attributes as surrogates for species-level patterns, including through satellite and aerial remote sensing, acoustic surveys, and image analysis (e.g., MODIS, Fig. 1). Nevertheless, surrogates frequently show poor congruency between different biodiversity elements (Mellin et al 2011, Rodrigues & Brooks 2007), and much variation remains concealed within mappable habitat features, yet innovative tools are available for censusing individual species and communities. For example, field survey data can now be integrated with high-resolution remote-sensing techniques to generate fine-resolution predictive maps, at least for shallow environments (Leaper et al 2012). Such biodiversity maps provide a necessary foundational layer for integrated coastal zone planning.

We here review new approaches for the collection of marine ecological data, with emphasis on species-level data that encompass regional to global scales. We also discuss analysis and interpretation of these and existing data to better inform management, with the ultimate goal of reduced anthropogenic impacts and improved environmental outcomes (Fig. 2).

Management need for big ecological data

As marine ecology evolves to understand the importance of global-scale processes and the fact that local studies may simply be unable to resolve meaningful patterns, so too has marine management embraced the multi-scale paradigm of ecosystem processes. Global-scale data facilitates management, in part through an improved awareness of how broader processes tie into observed local-scale trends, and also through lessons learned elsewhere. Threats to marine biodiversity are typically interactive and non-linear, consequently an empirical understanding of these threats requires replicated systematically-collected data from hundreds to thousands of sites. Such a span of sites is rarely available within a single jurisdiction, but instead requires multi-national coverage.

For example, marine protected areas represent the most reliable set of reference sites against which to assess fishing impacts. Yet too few marine protected areas (MPAs) exist within any individual country to allow analytical separation of interacting influences associated with key MPA design features (e.g. size, multi-zoned or single zone, proximity to other MPAs, fishing permitted within and nearby, level of compliance required, age). By contrast, influences of the different features can potentially be disentangled through coordinated global analyses of tens to hundreds of MPAs, thereby benefitting MPA planners worldwide (Edgar et al 2014).

Thus, detailed broad-scale information on changing patterns of marine biodiversity and associated threats are integral to improved management planning and assessment. The need for such data is particularly urgent in an era when threats to the health of the world's oceans—pollution, overfishing, habitat destruction, climate change, introduced pests —are universally recognised to be serious, pervasive and diverse (see, e.g., Halpern et al 2012, Halpern et al 2008, Jackson 2008). Moreover, the development of technology for new exploitative activities, such as deep-sea mining, translates to threats extending ubiquitously, including oceanic areas that were previously inaccessible. The existence of regional- and global-scale ecological data can greatly assist management in a variety of areas, including:

- Assessment and reporting of ecological condition through 'state of the environment' indicators, including progress of countries towards achieving international targets agreed under the Convention of Biological Diversity (Jones et al 2011).
- Identification of data gaps and management research priorities (Agardy et al 2011).
- Tracking of effects of changing climate (e.g., Bates et al 2014).
- Improved understanding of impacts of climate change on biodiversity heritage and ecosystem services (e.g., Graham et al 2015).

- Tracking of the spread of invasive species and better understanding impacts (e.g., Ruttenberg et al 2012).
- Improved understanding of impacts of fishing, including ramifications through food webs (e.g., Edgar et al 2011).
- Mapping distribution of biota for integrated coastal planning, including optimization of the location of marine protected areas and key biodiversity areas within a matrix of other zones (Edgar et al 2008, Leaper et al 2012).
- Provision of contextual information during assessment of the likely impact of local-scale developments on conservation values, including the level of irreplaceability of impacted zones (Raiter et al 2014).
- Assessment of local-scale impacts of oil spills, typhoons and other broad-scale threatening processes that act stochastically, through comparisons of impacted versus pre-impacted and reference sites (e.g., Edgar & Barrett 2000).
- Identification of threatened species and tracking trends in population recovery and decline (Richards 2014).

COMPILATION OF GLOBAL ECOLOGICAL DATA

Historical approaches to compilation of marine ecological data

The intimate association between human civilisation and the sea has resulted in a long history of observation of our marine environment and its biological inhabitants. This has been largely motivated by social and economic need. For instance, seafood has constituted an important element of global human diets for centuries, with the numbers and types of fish brought to market at various ports regularly documented. Many of these statistics have been collated by national and international organisations (e.g. Food and Agriculture Organisation: http://www.fao.org;

International Council for the Exploration of the Sea: http://www.ices.dk), and have subsequently been applied in studies of long-term changes in marine ecosystems (e.g., Callaway et al 2007, Engelhard et al 2014, Pauly et al 2005).

Many countries have also invested heavily in formal stock assessments, annual scientific surveys, and other long-term monitoring programmes aimed at understanding the dynamics of commercially-important fish species as well as their competitors, predators, and prey (e.g. Ricard et al 2012, Richardson et al 2006, Simpson et al 2011). Equally important, however, much marine biological exploration has been borne out of simple curiosity. For instance, the Challenger Expedition of 1873-76, which set out to document life on the deep sea bed, was a voyage of pure discovery with the very broad zoological remit to document "…the nature and distribution of the fauna of the ocean basins, and the form under which life was maintained under different physical conditions" (Thomson 1880). Countless other natural historical investigations, fisheries surveys, and research projects of varied geographic scope and temporal extent have additionally left an enormous legacy of observations on the distribution and variety of life in our seas. This data legacy contains crucial information regarding the past status of marine life, and its interactions with people and climate (Roemmich et al 2012). Nevertheless, the preponderance of presence-only observational data, and the relative scarcity of systematic surveys, mean that innovative methods are needed to fully exploit these data sets. Perhaps even more limiting, at least initially, is the fact that historical data have been scattered across multiple institutions and stored in many different formats (few of them digital), making it difficult for the research and conservation community to access and analyse this invaluable record of life in our seas.

Quantitative global scale data as a basis for marine ecological monitoring

Biodiversity data for marine systems now lie in online data meta-repositories developed over decades and involving global collaboration (Supplementary Table 1). Species lists (i.e., presence data: Fig. 1) generated, for instance, by the Ocean Biogeographic Information System (OBIS; http://www.iobis.org, and see OBIS case study box) and the Global Biodiversity Information Facility (GBIF; http://www.gbif.org) are invaluable for identifying biogeographic patterns but are much less useful for tracking change. Global effort towards enumerating species presence will inevitably vary through time in any set of amalgamated *ad hoc* surveys. Available ecological datasets in the marine realm have been conducted for a variety for different reasons through application of a variety of different methodologies, leading to sources of bias that may be difficult to pinpoint and thus account for in analyses.

Yet understanding the distribution and magnitude of ecological change is an urgent priority for human society, with management responses clearly hampered by the massive current shortfall of relevant monitoring information, high costs for targeted field surveys, and limited access to deep-water systems (Richardson & Poloczanska 2008). Nevertheless, following the lead from oceanographers, where international agreement on deployment of Argo floats has revolutionised their science (Durack & Wijffels 2010, Hosoda et al 2008, Roemmich et al 2009), standardised quantitative methodologies for monitoring marine biodiversity are increasingly applied over large scales, including efforts to coordinate among cabled observatories (e.g., International Ocean Network: http://msg.whoi.edu/ION).

A model example of cross-institutional collaboration in marine biodiversity observation involves deployment of the continuous plankton recorder (Fig. 1), with consistent species-level data on plankton densities now routinely obtained through ships of convenience travelling in European, North American, Australian and Antarctic seas (Fort et al 2012, Hosie et al 2003, Richardson et al 2006). Multi-national networking and data sharing between animal tracking groups, most notably the Global Tagging of Pelagic Predators program (TOPP; http://www.gtopp.org/), comprises another key example of accelerating scientific understanding gained through collaboration (Bessudo et al 2011, Block et al 2011), where integrated outputs equal more than the sum of parts.

Citizen science possesses arguably the greatest potential for scaling up biodiversity monitoring globally, to achieve coordinated species-level observations across scales otherwise impossible to cover because of impractical cost for professional research teams. Most citizen science activities will, however, be confined to species and systems where relatively simple observations can be made, particularly for semi-aquatic species such as birds, mammals and intertidal dwellers, and also for macroscopic organisms living in shallow subtidal habitats accessible by divers.

Coordination of species-level observations across regional scales has been pioneered largely by amateur bird-watchers, with leadership from Birdlife International and its national partners (BirdLife International 2004a). Outputs of these programs not only possess major scientific significance, but have also catalysed coordinated conservation planning and management at continental to global scales (Eken et al 2004, Langhammer et al 2007, Stattersfield et al 1998). While most of this effort is directed at terrestrial habitats, huge datasets are also accumulating for shore- and seabird species, including ocean wanderers (BirdLife International 2004b).

The number of citizen science initiatives involving divers is also accelerating rapidly, with Reef Check (www.reefcheck.org/*)* (Hodgson 1999), Reef Environmental Education Foundation (www.reef.org/*)* (Francisco-Ramos & Arias-González 2013), and Reef Life Survey (www.reeflifesurvey.com/) (Edgar & Stuart-Smith 2014), in particular, all gathering standardised biodiversity observations and extending globally in reach. Compromises are required by citizen science organisations when setting the balance between volunteer engagement and complexity of methods (Holt et al 2013). At one end of the spectrum are organisations with primary focus on public participation and educational outputs, which therefore employ simple methods that are manageable by all. At the other extreme, Reef Life Survey concentrates on quality of scientific outputs, training a small set of enthusiastic divers to a scientific level in underwater visual census techniques, but at the cost of wide engagement (see RLS case study box).

A case study in compilation of diverse observational data: OBIS

The lack of a sufficient system for the retrieval of marine biological data was recognised at the outset of the decade long Census of Marine Life (CoML) (Grassle 2000). One of the major outcomes of the CoML was the creation and maintenance of a central, standardised and open access portal, the Ocean Biogeographic Information System (OBIS; www.iobis.org). OBIS has since established collaborative links with the Global Biodiversity Information Facility (GBIF) and developed into the largest primary provider of spatial records of marine biodiversity data at a global scale, with ~42 million geo-referenced records of the occurrence of marine taxa (February 2015, Fig. 3), >75% of them resolved to species level or better.

OBIS has become an invaluable source of marine biodiversity data for research, used extensively (>900 citations listed in Google Scholar) to assess, for example, the degree to which various environmental factors predict global marine biodiversity and the spatial congruence between biodiversity and human impacts (Tittensor et al 2010). OBIS has also been used to validate models of marine species distributions (Ready et al 2010), and to parameterise models of predicted range shifts of exploited fish and invertebrate species (Jones & Cheung 2014). Further, analysis of OBIS data can help to identify knowledge gaps, documenting taxonomic and spatial biases in data availability (Appeltans et al in press, Miloslavich et al in press), which reveal chronic under sampling in Earth's largest biome, the deep pelagic ocean (Webb et al 2010).

Despite these high levels of use, the full potential of OBIS has yet to be realised. For instance, OBIS is a significant repository of historical data, with an average of 1,800 observations daily since the 1960s (Appeltans et al in press). Some analyses have used this temporal dimension to document trends in component datasets within OBIS that focus on specific taxa or communities (e.g., Dornelas et al 2014), but comprehensive analyses of changes in biodiversity through time at regional and global scales have yet to be attempted. OBIS data may also contribute to specific objectives in marine spatial planning (Caldow et al 2015), and we expect such management and policy applications to increase with the maturation of the science of biodiversity informatics (Costello & Vanden Berghe 2006, Hardisty et al 2013).

The strengths of OBIS include its size and its taxonomic and geographical breadth, with occurrence records for ~160,000 marine taxa in all major marine regions of the world. Importantly, all data are freely accessible through the existing web portal, with advanced access possible through the underlying PostgreSQL database and web services in development. However, OBIS is an amalgam of >1,700 separate datasets, gathered by different research groups with varying goals and methodologies. The result is an unstructured database with variable data quality. OBIS records require geographic and taxonomic information, with additional variables such as date and depth of recording, environmental variables, and method of data collection sometimes available, although often some or all of these are lacking. Values also vary in precision, with clear errors in distribution records reported (e.g., Robertson 2008). Some of these problems are addressed by the quality control procedures employed when adding data to OBIS (www.iobis.org/node/47), and the additional tests recently developed and implemented to automatically quality control marine biogeographic databases including OBIS (Vandepitte et al 2015). This will go some way to ensure data are adequately structured and complete where required, however the onus remains on the user to perform appropriate data control and manipulations when developing a dataset suitable for any specific analysis.

In addition to the unstructured nature of the data and the spatial and taxonomic biases present in OBIS, another important issue is that of imperfect detectability, i.e., the ability to separate true absence of a species from an area from instances where it was present but undetected. Large-scale analyses often assume that the absence of a species in the data represents a true absence (i.e., that its detection probability is 1, Kéry et al 2010, Monk 2014). Clearly, this is rarely the case. In marine data, detection probabilities are likely to be substantially <1 as a consequence of the logistical challenges of surveying (Bates et al. 2014). Moreover, because sampling is not uniformly effective, catchability varies among individuals and species, as well as over space and time (see, e.g., Fraser et al 2007, Royle et al 2007). Multiple statistical methods exist to account for imperfect detection (see Table 1; Bird et al 2014), and applying them to OBIS data – where the contribution of numerous datasets likely means a large effect of detection bias in the final database – is likely to lead to substantial increases in the utility of this vast database to marine conservation science.

A case study in compilation of systematic quantitative data: Reef Life Survey

The Reef Life Survey program (RLS) represents a large-scale standardised approach to surveying and monitoring marine biodiversity through the engagement of committed recreational SCUBA divers (Fig. 4, Edgar et al 2014). It provides a structured framework in which trained recreational divers provide observations that allow tracking of changes in subtidal habitats for scientific and management applications. RLS is based on a model that primarily emphasises generation of large quantities of scientific-quality data, rather than the broad public outreach central to most marine citizen science programs.

RLS evolved from a pilot project funded by the Australian Government through its Commonwealth Environmental Research Facilities program from 2007-2010, which saw the training of an initial team of divers, and assessment of the suitability of methods, data quality, and cost-effectiveness of the approach. Survey methods were based on visual census techniques applied over two decades by University of Tasmania and tropical eastern Pacific researchers in MPA monitoring studies (Barrett et al 2009, Edgar et al 2011, Edgar & Barrett 1999). The RLS model successfully allowed data collection over greater geographic and temporal scales than possible by professional scientific teams, without sacrificing taxonomic resolution and other detail. Following the pilot project, the not-for-profit Reef Life Survey Foundation (http://www.reeflifesurvey.com/) was formed to train committed divers in systematic underwater visual census surveys, refine data entry procedures, and operate ongoing field activities through a combination of targeted field campaigns and *ad-hoc* surveys of local and vacation sites by trained divers.

RLS methods cover four major components of biodiversity along 50-m long transect lines set on subtidal rocky and coral reefs – fishes, large mobile macro-invertebrates, sessile invertebrates, and macro-algae. The abundances and sizes of all fish species sighted within 5 m x 50 m belts either side of the transect line are recorded by divers, and the number of all mobile invertebrates (echinoderms, crustaceans and gastropods) >2.5 cm in length, and cryptic fish species, are counted within narrower 1 m x 50 m belts during close searches of the substrate. Digital photoquadrats are also taken every 2.5 m along the transect lines for later estimation of the cover of sessile invertebrates, macrophytes and abiotic habitat types using appropriate software (e.g. Coral Point Count, Kohler & Gill 2006).

Individual training is provided to selected recreational divers, following initial screening for diving experience and commitment. Trainees follow experienced divers along transect blocks, duplicating surveys, then receive feedback on elements requiring improvement until data collected by trainees closely match those of the trainer (Edgar et al 2014). An analysis of data quality showed that for data collected on the same dives, the variation between newly trained divers and experienced scientists was negligible (<1%) in comparison to differences between sites and regions (Edgar & Stuart-Smith 2009). In addition, a degree of self-regulation of data quality was observed where initial data quality upon completion of training was positively related to ongoing involvement and productivity of RLS divers following their training (Edgar & Stuart-Smith 2009). Thus, the most enthusiastic recreational divers tend to also collect the most accurate data, participate more frequently, and stay involved in survey activities for longer. More than 100 active RLS divers participate at present, and standardised, quantitative data have been collected at >2,500 sites, including >500,000 abundance records for >4,500 species. Many sites have been surveyed on multiple occasions, in some cases annually each year since 2007, and these numbers continue to grow.

The global RLS data not only provide broad context to local surveys and monitoring data, but also allow exploration of patterns in biodiversity where species-level resolution over large scales is needed (e.g., Stuart-Smith et al 2013), or that have not been possible previously due to insufficient commonality when amalgamating multiple datasets collected using multiple methods. In addition, the standardised nature of RLS data facilitates the analysis of data over large spatial scales by removing some observer-related sampling biases that may be present in citizen science programs with less stringent training requirements.

In a global assessment of ecological differences between MPAs and fished locations, Edgar et al. (2014) were able to include an order of magnitude more MPAs than any previous study based on standardised data, allowing quantitative comparisons not achievable through approaches such as meta-analyses. They found that fish communities in most of 87 MPAs investigated were indistinguishable from reference fished communities and so largely ineffective; however, MPAs characterised by no-fishing regulations, high levels of enforcement, established more than 10 years, large in area, and isolated from fished areas by habitat boundaries, were extremely effective, with substantially elevated biomass of large fishes.

The continually expanding RLS dataset should prove invaluable as a baseline for assessing changes in global shallow-water marine biodiversity associated with accumulating and expanding threats, and for allowing the identification and tracking of the trajectories of threatened species populations. In the field of macroecology, it allows investigations on how whole communities of mobile macroscopic organisms from multiple phyla (Chordata, Echinodermata, Arthropods, Mollusca) and classes interact at landscape levels (Webb 2012). The terrestrial analogue would be a global dataset that combines quantitative co-located surveys of mammals, birds, reptiles, amphibians, insects, spiders, myriapods and gastropods. The RLS dataset also provides unique opportunities for tracking international marine conservation targets (Group on Earth Observations Biodiversity Observation Network 2011) and the effectiveness of national and international policy, as well as local and regional management strategies.

Looking forward: global-scale manipulative experiments

While accurate mapping of patterns of biodiversity at regional to global scales represents a huge advance in ecological knowledge, it will inevitably lead to more questions than answers about underlying causes. Nevertheless, embedded within broad-scale spatial and temporal datasets is much information about ecological process. This can be inferred, in part, through ‘natural experiments', where factors of interest exhibit gradients; however, causality can rarely be attributed in such studies because of the range of environmental covariates that are typically also inter-correlated with the primary factor.

Thus the coordination of experimental networks represent an emerging class of broad-scale investigation, where controlled manipulative experiments are replicated in different regions. Such networks typically include researchers from a variety of countries, all agreeing to apply the same experimental or mensurative survey protocol in order to combine the rigor of experiments with observational data on environmental gradients that cannot be manipulated.

An example of this approach is provided by the Zostera Experimental Network (ZEN, www.zenscience.org), which seeks to understand how complex regional and local processes interact to affect community and ecosystem structure in eelgrass (*Zostera marina*) beds (Fig. 1). In one experimental ZEN study where standardised nutrient addition and grazer reduction treatments were applied at 15 locations in 7 countries, algal biomass was generally found to be locally controlled more by top-down (grazing) rather than bottom up (nutrient fertilization) processes, but with global-scale patterns of biodiversity (grazer and eelgrass richness) strongly influencing local-scale outcomes (Duffy et al 2015).

Global coordination has the potential to address some of the greatest conservation challenges of our time. In particular, management interventions repeated across the seascape, such as MPAs, comprise a related class of observational experiment. Each intervention represents a deliberate manipulation of the natural environment, consequently the sum of such interventions is a direct analogue of the classical ecological experiment with replication, but one that usefully has been conducted at an appropriate scale for improved management understanding (Walters & Holling 1990).

Studies of human manipulations are, however, often confounded when the management action is not randomly distributed across the seascape but is located in a preferred combination of environmental conditions, such as fish farms in deep sheltered embayments, sewage outfalls on headlands with good current flow, and MPAs in locations with few exploitable living resources. In such situations, reference sites should be selected to match environmental conditions at the intervention sites, or environmental conditions at the set of reference and intervention sites quantified, and their effects then modelled and partitioned separately from the reference versus intervention comparison.

Moreover, management actions at different locations are rarely identical; rather, they are generally modified to suit the particular set of local circumstances. Such modifications add statistical noise to analyses if the intervention is categorically regarded as either present or absent; however, when local information is explicitly recognised and modelled, this variability often becomes of prime scientific interest amongst analytical outcomes. For example, the fact that fish biomass recovers within MPAs is now well-recognised, consequently scientific and public interest is more directed towards understanding how this recovery is affected by such factors as MPA size, boundary configuration, governance framework, level and type of community support, and policing (Daw et al 2011). Similarly, important questions related to fish farms and sewage outfalls include how environmental impacts are modified by type of treatment, current flow and nutrient loads.

Viewed from a satellite, broad-scale experimental networks highlight the pseudoreplication involved when outcomes of small-scale ecological experiments are generalised widely, given that experimental plots are traditionally dispersed over a local area of <10 km span, and coalesce into a single point when viewed from high altitude. Progress in ecology clearly depends on improved general understanding derived from experiments with worldwide span. Coordinated experimental networks clearly have potential to greatly assist this process by providing the maximum possible generality in inference. Nevertheless, to achieve this, replication needs to be sufficiently large to overwhelm statistical noise introduced by idiosyncratic local-scale environmental variability at each site investigated.

ANALYSIS OF GLOBAL ECOLOGICAL DATA

Integrating varied data types and sources

Conservation science demands strategies that are coherent, trans-disciplinary and integrated, with access to data that tracks meaningful patterns and trends (Pressey et al 2007). Socio-economic frameworks are also required for statistical and conceptual models, with input of data spanning spatial scales, trophic levels and organism lifespans (Bestelmeyer et al 2011). Consequently, ecological datasets alone are insufficient for answering many of the major outstanding questions in conservation science (Sutherland et al 2009), but need to be combined with covariate data from physical, social and economic domains (see Supplementary Table 1 for relevant online datasets).

Ecological data relevant to population dynamics still primarily rely on visual census methods for abundance and taxonomic accuracy, regardless of significant developments in the field of automated detection of biotic components (Mallet & Pelletier 2014). Despite increasing operational costs, *in situ* visual survey methods, such as RLS (see RLS case study box), the Australian Institute of Marine Science Long Term Monitoring Program (Sweatman et al 2011), and the U.S. National Park Service Kelp Forest Monitoring Program (Rogers-Bennett et al 2002), tend to be best established because of their wide utility. They are non-destructive and can incorporate multiple methods focused on a wide range of species, such as vertebrate and macro-invertebrate observations, plus detailed image analyses.

Supplementary to these programs are destructive methods such as scientific benthic trawls and netting that have a longer history than SCUBA-based approaches. For deep and visually-constrained areas, destructive methods are the most cost effective option (McKenzie et al 2012, Varjopuro et al 2014). Videos and cameras designed to capture visual data without the complication of human submersion complement these approaches (Clarke et al 2012, Mallet et al 2014).

Surveys of organisms such as corals or fish schools can be aided by the use of multibeam sonar (Zieger et al 2009), while automated sampling of surface waters through the ships of opportunity program tend to focus on planktonic densities (Williams et al 2006). While raw data continue to remain with the collecting institutions, an imperative exists for the data to be standardised in terms of units of measurement (species identity, metric sizes and weights, abundance bin classes, site geo-positioning accuracy and coordinate system). Online portals of taxonomic information such as FishBase (http://www.fishbase.org/), IUCN (http://www.iucnredlist.org/), WoRMS (http://www.marinespecies.org/) and Corals of the World (http://coral.aims.gov.au/) permit the standardised description of species where taxonomic resolution allows (Supplementary Table 1). These portals also include, or are developing, attribute information necessary for many size-based trophic models and conservation vulnerability estimates, although data on attributes such as biological traits and conservation status are lacking for the majority of marine species (Tyler et al 2012).

The rapid growth of sophisticated remote sensing tools and analytical techniques has enabled the provision of global datasets describing the physical, biological and chemical environments for the surface waters (Andréfouët & Hochberg 2005, Collin et al 2012). Satellites that polar circumnavigate the planet provide daily recordings of radiation and reflectance detailing changes in ocean colour, temperature, light attenuation, wave heights, flood plumes, ice coverage, storm activity and cloud cover (Chassot et al 2010, Devlin et al 2012, Tyberghein et al 2012, Young et al 2011). Secondary products derived from these observations using sophisticated algorithms include models of isolation from disturbance, circulation patterns, tidal dynamics, chlorophyll concentrations, nutrient concentrations, temperature anomalies and oxygen saturation levels (Basher et al 2014). Extrapolation of these models to past and future time periods is also available for marine modellers seeking to implement scenario-based predictions.

Dynamic prediction systems based on these data now present global 'real-time' measurements of marine dynamics such as vulnerability to thermal stress (e.g., NOAA Coral Reef Watch, Liu et al 2012). The spatial resolution is a compromise between image extent, temporal cycle and radiation attributes rather than specifically aligned to ecological scales. While agencies such as NASA continue to provide data online, the emergence of collected datasets in web portals has stimulated marine research: Bio-Oracle (Tyberghein et al. 2012) and GMED (Basher et al. 2014) are two leading examples. Supportive information on the physical structures of the marine environment such as coastlines, bathymetry, infrastructure, political boundaries, activity zoning boundaries and shipping routes are now easily downloadable at fine resolution (see Reefs at Risk; http://www.wri.org/resources/data-sets/reefs-risk-revisited), with research ongoing to make climate and geographic data more accessible, such as through the FetchClimate portal (http://research.microsoft.com/en-us/projects/fetchclimate/).

High temporal intensity, high spatial definition data for specific programs is obtained through in situ loggers/sensor networks (Hendee et al 2012, Kininmonth 2007, Marin-Perianu et al 2008). Argos, with their global network of ocean drifting buoys and mobile sensors, is certainly the most comprehensive, although other national infrastructure initiatives such as the Australian Integrated Marine Observation System (http://www.imos.org.au) are noteworthy. Sensor data is expensive primarily due to the maintenance schedule in the harsh marine environment, but the capacity to record precision data at depth has unrivalled value, especially in the oceanographic modelling arena (Bondarenko et al 2010).

Marine data specific to human impacts are disjointed and sparse, often reflecting the national interest rather than contributing to a global repository. Economic data are varied in quality across the globe, including in the fisheries industry, despite their high importance (Bodin & Österblom 2013, Folke 2015). Data required to compare basic human extraction practises, such as proportion of fish traded nationally or internationally (Cinner et al 2013), are not consistently available. International corruption indices (e.g., Transparency International; http://www.transparency.org/) are available for national scale analyses dealing with governance effectiveness, but associating this social measure to ecological processes is difficult due to the scale mismatch.

However, many indicators of human density, activity and wellbeing can be captured irrespective of national priorities. In particular, the application of remotely-sensed images to measure light intensity as a surrogate of industrial activity and population density has helped to determine relative impacts of coastal developments (Pesaresi et al 2013). Transport activity in the marine environment can also be estimated through shipping vectors, although fishing fleet activity is more difficult to remotely observe and requires the use of vessel monitoring systems (Gerritsen & Lordan 2011, Lee et al 2010). Population census data defining the density of people is more robust and available at a finer scale, such as 30 arc-second grid cells for GPWv4 (Center for International Earth Science Information Network 2014) and 100 m cells for Worldpop (www.worldpop.org). This also includes indices of human poverty (Stevens et al 2015), albeit lacking in functional attributes, such as the number of fishers.

Capacity to define the functional aspects of human society is presented by demographic health surveys (www.dhsprogram.com), which obtain data through dense, nationally-representative household surveys across a range of topics. Social boundaries such as Exclusive Economic Zones are defined but seldom have rigid behavioral impacts. Even well-defined marine protected areas have issues with enforcement (Edgar et al. 2014). International policy agreements (ECOLEX; www.ecolex.org) are available for interrogation but down scaling these documents to match ecological processes is conceptually difficult (Treml et al 2015). Developing large-scale data on social activities and interactions has seen the collection of communication data from mobile phones (Deville et al 2014), internet software (e.g., Facebook), bank transactions, credit card usage and money transactions (Barabási 2005, Song et al 2010), yet the specific applications to the marine environment remain sparse. Similarly, many industrial activities in the marine environment (such as wind farms, fish farms and oil platforms) require rigid environmental monitoring in order to fulfill operational licensing requirements, but the data remain in the private domain and are rarely cross-referenced.

To counteract this data paucity in the face of increasing levels of extraction and destruction (Halpern et al 2008), conservation efforts have attempted to specifically identify key processes that contribute to the decline in marine health and spatially describe them. Reefs at Risk (Burke et al 2011) and Status of the Coral Reefs (Wilkinson 2008) supply damage estimates with contributing factors. In recent years the focus on regime shifts has spawned online databases seeking to collate case studies (e.g., Resilience Alliance and Santa Fe Institute 2004). While these meta-data portals provide a start, much greater communication, coordination and data provisioning is needed across disciplines to address global change challenges.

New approaches to statistical analysis of big ecological data

Analytical approaches to emerging global marine datasets will need to accommodate two main data challenges. First, the data will be increasingly complex, requiring novel statistical solutions. Second, the size of many new datasets will lead to storage and computational challenges. Here we outline some of these challenges as they relate to the analysis of marine ecological datasets.

Most early statistical analyses were designed under the assumption that all samples were collected under similar conditions and inference was aimed at determining the effect of one or two factors. However, this scenario is increasingly unlikely in datasets that cover large areas, have many distinct observers, occur over long time periods, or have many predictor variables. Furthermore, distinct sampling units may have differing sampling conditions that influence the quality, quantity or error structure of data. If not accounted for, these differences can lead to significant bias and erroneous conclusions (Diniz - Filho et al 2003, Kühn 2007).

Many of the issues related to sampling bias can be addressed using hierarchical models, an umbrella term for a class of parametric analyses in which model parameters are themselves considered to be drawn from some probability distribution (Wikle 2003). Mixed-effects models are gaining popularity in ecological research (Bolker et al 2009), while in global datasets, the use of metadata such as location, sampling conditions and survey team experience can be used to partition and account for sources of variability (Bird et al 2014). In addition, hierarchical models are extremely useful in a meta-analytical context, where data from large numbers of independent studies can be integrated into a unified analysis.

Ecologists are now more acutely aware that samples taken closer together are likely to be more similar than those taken farther apart (Bivand 2014, Cliff & Ord 1968), and a wide range of approaches have been developed to address this issue (e.g., Dormann et al 2007) (Table 1). Perhaps most significantly, user-friendly and open source statistical packages are making such analyses more accessible. Bayesian models can be used to model complex spatio-temporal dependencies within the data using conditional likelihoods, resulting in models that better reflect the ecological processes of interest. However the Bayesian approach is not often applied to large-scale data analyses due to the computational burden of Monte-Carlo Markov Chain methods used for inference. A more efficient approach to Bayesian inference is Integrated Nested Laplace Approximation (INLA, Table 1). INLA uses approximate inference to arrive at the posterior distributions that would normally be inferred using MCMC, but arrives at a solution much more quickly (Raudenbush et al 2000).

Another important source of bias in abundance or occurrence datasets is failed detection (see OBIS box), where recorded absences of species are partly due to sampling error (Tyre et al 2003). Approaches to correcting for this bias rely on a modified sampling procedure, in which replicate observations are used to separately model the process of interest (such as presence/absence) and the probability of accurately detecting a species given that it is present. By estimating the rate of detection, the overall probability of occurrence is adjusted accordingly (MacKenzie et al 2002).

Occupancy detection models belong to a class of state-space models (SSMs), in which observations are assumed to be dependent on some underlying state. Conceptually, this idea can be extended to a wide range of scenarios, allowing for modelling sampling-related biases such as misidentification of species, uncertainty in location, or variation in life-history stages or other ecological processes (e.g., Borchers & Efford 2008). SSMs are increasingly used in a Bayesian context, as they allow complex hierarchical models for ecological processes of interest (e.g., King 2012). A challenge with SSMs is that they can be computationally intensive, given that analyses must impute values for all hidden parts of the model (see Table 1).

Machine learning (ML) approaches offer an alternative to parametric models. ML has the advantage that it does not rely on distributional assumptions in order to make predictions, and has been used to identify global-scale biodiversity patterns from gridded raster datasets and geo-located survey data (e.g., Stuart-Smith et al 2013). More recent ML approaches such as Quantile Regression Forests

(Meinshausen 2006) and Boosted Regression Trees (Elith et al 2008) can provide confidence intervals around predictions, or infer linear relationships between variables and covariates. Because they do not rely on distributional assumptions, Random Forest approaches have been used with re-weighting procedures to convert available low-resolution areal data to predictions at high resolutions (Deville et al 2014, Leyk et al 2013). In the context of large and complex datasets, ML approaches can suffer as the number of model nodes increases exponentially with the number of observations, rapidly overwhelming the capacity of many computing systems. A more recent development is the use of decision jungles, which modify traditional random forests using a probabilistic method of merging nodes in a directed acyclic graph (Shotton et al 2013).

Another class of ML model is Gaussian Process Models (GPMs). GPMs are essentially a smoothing technique in which the response data are modelled as the outcome of some multivariate Gaussian process - any set of functions with a joint Gaussian distribution and zero mean (Rasmussen 2006). GPMs can be fit to data in multidimensional space and then used to construct Bayesian priors for expected response values in unsampled space (Banerjee et al 2008).

However in all of the above applications, the size of databases – both for observation data and predictor covariates – is increasingly a limiting factor. In particular, datasets generated by electronic tagging (e.g., Block et al 2011), video (e.g., cabled observatories: Matabos et al 2014), acoustic recorders (Korneliussen & Ona 2002), or environmental monitoring packages designed to measure multiple physical parameters (such as temperature or fluorescence), can rapidly grow to terabytes of data. In many analyses, high performance computing (HPC) clusters and cloud computing provide solutions by allowing large prediction problems to be split into many small tasks through model fitting of subsets of the prediction dataset. Platforms such as Microsoft Azure, Google Compute Engine, and Amazon Elastic Cloud Compute are all suited to this kind of task, which has been described as ‘embarrassingly parallel’ (Wilkinson & Allen 1999), meaning the data and analyses can be subset into smaller independent packages without influencing the result.

Where analysis of the complete dataset is limited by memory, map-reduce algorithms such as Hadoop split data according to some criteria prior to performing analyses separately on data subsets on separate cores. The results are then combined. Many kinds of problems can be approached in this way, particularly where data first need to be classified or sorted, then aggregated using some calculation. In these cases, algorithmic analyses such as data mining are useful for isolating particular patterns in the data.

Alternatively, where analysis of the dataset as a whole is required, distributed computing splits the analysis task between cores of a HPC cluster as in an embarrassingly parallel computing environment, with the difference that the nodes in the cluster use a message passing interface in order to allow different parts of the algorithm to interact with one another. A simple example of distributed computing in an environmental context might be the aggregation or down-sampling of ecological data from a large network of sensors (Porter et al 2012). Finally, some researchers are turning to crowdsourcing of analytical tasks by asking internet-based volunteers to perform image processing tasks that can be done easily by people but are complex for computers (Shamir et al 2014).

As ecological research becomes increasingly data intensive, and involves crowdsourcing of data processing (e.g., digitising raw data), we will be challenged to maintain the integrity of our workflows; data acquired and integrated across multiple scales will require multiple data formats, statistical approaches and software packages (Levy et al 2014). Staying on top of such diverse sources of information and their respective complexities may require a unified framework for analysis, allowing greater reproducibility of research as well as iterative learning when new data become available (Michener & Jones 2012). In addition, the availability of data is often now outpacing its usability by managers, with many of the datasets described here requiring significant technical expertise to access and use. In this case, simulation programs such as Ecopath (Pauly et al 2000) and systematic conservation planning tools such as MARXAN and C-Plan (Carwardine et al 2007) can be invaluable for distilling dense ecological data into actionable management goals.

Trend indicators for marine biodiversity and associated threats

Tracking trends in marine biodiversity in relation to national and international biodiversity conservation targets requires the summary of complex ecological responses to anthropogenic threats within 'state' indicators (Jones et al 2011, Smale et al 2011). Thus, information that is multispecies and multidimensional needs to be reduced to comprehensible units that can be mapped or graphed. In order to best guide policy and management, indicators need to not only be understandable to the public and policy-makers, but also respond to particular threats in a predictable manner, allowing assessment of the success of mitigation efforts directed at that threat (Collen & Nicholson 2014).

Despite decades of research focussed on the development and selection of indicators for this purpose (e.g., Fulton et al 2005, Rice 2000), challenges remain when (a) balancing the need for comparability across large scales without losing substantial ecological detail (Pereira et al 2013), (b) establishing empirical links between indicator values and threats (Collen & Nicholson 2014), and (c) quantifying the specificity of indicators to threats (Link et al 2010). Broad-scale indicators of ecosystem state which respond exclusively to a particular threat may not exist, but identifying those that are most responsive to each threat and that can quantify and account for interactions among threats remains an important research goal (Nicholson et al 2012).

The investigation of broad-scale indicators for marine biodiversity has disproportionately focussed on ecosystem responses to commercial fishing, where management targets are often clearly defined and data most readily available. Substantial empirical and theoretical support exists for indicators based on the size or biomass spectrum of the whole fish community (Graham et al 2005, Jennings & Dulvy 2005), for example. However, spatial variation in the importance of environmental drivers and community structure in determining size spectra have not been evaluated at the global scale, and gear selectivity and methodological inconsistencies have imposed substantial barriers to broad-scale application and interpretation (Shin et al 2005). In addition, responses in community size spectra to changes in fishing pressure have been suggested to be too slow to direct fisheries management responses, which often occur on a year-to-year basis (Nicholson & Jennings 2004). Thus, size-spectra are arguably most useful for guidance on medium-term policy (Jennings & Dulvy 2005).

Broad-scale indicators for threats associated with climate change, pollution and invasive species have been more poorly studied, with little guidance on sensitive indicators, let alone understanding of specificity or how response times compare with timeframes for management options. Current management and policy targets are also necessarily vague for such threats, and understanding of their relative importance is currently based primarily on expert opinion (Halpern et al 2007). Refining these targets with explicit quantitative goals is largely dependent on the availability of informative indicators to measure progress against these targets. Likewise, prioritising management and directing policy is dependent on research which incorporates threat indicators and can identify critical interactions, non-linearities, and links between threats and ecosystem functions and services.

CONCLUSION

Our living marine heritage is declining at an accelerating rate (McCauley et al 2015), in part because changing ecological patterns occur below the sea surface, and so are largely invisible to the public, including scientists, managers and policy makers. Even basic information, such as how well countries are complying with international commitments to the Convention on Biological Diversity, is essentially lacking. The general lack of systematic ecological monitoring data for tracking trends in marine condition, and scarcity of comprehensive analyses of existing data, contribute significantly to this situation.

Survey data are needed that can be used to describe patterns of biodiversity at regional- to global-scales, and that also link to fine-scale ecological data. Monitoring programs must accommodate study designs that are systematic, species-level, and spatially intelligent, while also anticipating the use of the new wave of ecological and geostatistical modelling approaches for analysis, including protocols that account for missing data. Conservation science needs a large-scale and long-term view of 1) data that are necessary for tracking responses and impacts, 2) fundamental metrics needed for reporting, and 3) ways to integrate past, current and future methods to provide both the fine-scale inference required for policy and intervention, and an integrated view of the global picture.

Prioritisation of conservation actions requires, amongst other inputs, maps of the world delineating the state and trends for various conservation-related metrics. The framework exists to achieve this goal by embracing the 'big data' nature of the problem, and incorporating a wide perspective into the design of studies. Important next steps include 1) recognition that biodiversity conservation depends fundamentally on persistence of species, and that monitoring trends in species population numbers is pivotal to conservation strategies, 2) consideration of spatiotemporal biases present in survey outputs, 3) augmentation of existing sampling protocols with new approaches to help overcome these biases, including expanded programs that leverage latent support from citizen scientists, 4) utilisation of large-scale covariate datasets, which are becoming increasingly available, 5) applying new methods of modelling that are facilitated by the massive recent increase in computational power, and 6) broadcasting ecological monitoring results in a way that makes them globally available and locally relevant.

FUTURE ISSUES

1. Conservation science and ecology will both greatly benefit from expanded global data-gathering and experimental networks, and new mechanisms for rapid retrieval and collation of marine biodiversity monitoring data.
2. Improved coordination and communication is needed among disciplines, so that physical, biological, economic and social data are available at matched and relevant spatial and temporal scales to address global environmental challenges.
3. Big data techniques used in other disciplines should be adopted more widely in conservation science, as well as new research collaborations established to develop tools for storing, managing, accessing, linking, visualizing and analysing data.
4. Biodiversity targets that are meaningful in the context of global change need to be identified, along with appropriate metrics that can be reliably used to track progress towards these targets.

DISCLOSURE STATEMENT

GJE and RS-S are Board Members of Reef Life Survey Foundation. The authors are not aware of any other affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

GJE and RDS-S were supported by the Australian National Environmental Research Program (NERP) Marine Biodiversity Hub and the Australian Research Council. TJW is a Royal Society University

Research Fellow. AHJ is supported by a NERC studentship through the ACCE Doctoral Training Partnership.

LITERATURE CITED

Agardy T, Di Sciara GN, Christie P. 2011. Mind the gap: addressing the shortcomings of marine protected areas through large scale marine spatial planning. *Marine Policy* 35: 226-32

Anderson DP, Cobb J, Korpela E, Lebofsky M, Werthimer D. 2002. SETI@ home: an experiment in public-resource computing. *Communications of the ACM* 45: 56-61

Andréfouët S, Hochberg E. 2005. Multi-scale remote sensing of coral reefs. In *Remote Sensing of Coastal Aquatic Environments*, ed. RL Miller, CED Castillo, BA Mckee, pp. 297-315: Springer Netherlands

Appeltans W, Dujardin F, Flavell M, Miloslavich P, Webb TJ. in press. Biodiversity Baselines. In *Global Environment Facility International Waters Transboundary Waters Assessment Programme (Open Oceans)*

Banerjee S, Gelfand AE, Finley AO, Sang H. 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70: 825-48

Barabási A-l. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435: 207-11

Barrett NS, Buxton CD, Edgar GJ. 2009. Changes in invertebrate and macroalgal populations in Tasmanian marine reserves in the decade following protection. *Journal of Experimental Marine Biology and Ecology* 370: 104-19

Basher Z, Bowden DA, Costello MJ. 2014. Global Marine Environmental Datasets (GMED). http://gmed.auckland.ac.nz

Bates AE, Barrett NS, Stuart-Smith RD, Holbrook NJ, Thompson PA, Edgar GJ. 2014. Resilience and signatures of tropicalization in protected reef fish communities. *Nature Climate Change* 4: 62-67

Besbeas P, Freeman SN, Morgan BJ, Catchpole EA. 2002. Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* 58: 540-47

Bessudo S, Soler GA, Klimley AP, Ketchum JT, Hearn A, Arauz R. 2011. Residency of the scalloped hammerhead shark (Sphyrna lewini) at Malpelo Island and evidence of migration to other islands in the Eastern Tropical Pacific. *Environmental Biology of Fishes* 91: 165-76

Bestelmeyer BT, Ellison AM, Fraser WR, Gorman KB, Holbrook SJ, et al. 2011. Analysis of abrupt transitions in ecological systems. *Ecosphere* 2: art129

Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, et al. 2014. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation* 173: 144-54

BirdLife International. 2004a. *State of the world's birds*. Cambridge, UK: BirdLife International

BirdLife International. 2004b. *Tracking ocean wanderers: the global distribution of albatrosses and petrels. Results from the Global Procellariform Tracking Workshop, 1-5 September, 2003, Gordon's bay, South Africa.* Cambridge, UK: BirdLife International

Bivand R. 2014. "The Problem of Spatial Autocorrelation:" forty years on.

Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, et al. 2011. Tracking apex marine predator movements in a dynamic ocean. *Nature* 475: 86-90

Bodin Ö, Österblom H. 2013. International fisheries regime effectiveness-Activities and resources of key actors in the Southern Ocean. *Global Environmental Change* 23: 948-56

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, et al. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution* 24: 127-35

Bondarenko O, Kingsford M, Kininmonth SJ. 2010. Deployment of Wireless Sensor Network to Study Oceanography of Coral Reefs. *International Journal On Advances in Networks and Services* 3: 85-95

Borcard D, Legendre P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153: 51-68

Borchers DL, Efford M. 2008. Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics* 64: 377-85

Brunsdon C, Fotheringham S, Charlton M. 1998. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)* 47: 431-43

Buckland ST, Anderson DR, Burnham KP, Laake JL. 1993. *Distance sampling: estimating abundance of biological populations*: Chapman & Hall

Burke L, Reytar K, Spalding MD, Perry A. 2011. Reefs at risk revisited. *World Resources Institute* 1: 1-112

Burnham KP, Anderson DR, Laake JL. 1980. Estimation of density from line transect sampling of biological populations. *Wildlife monographs*: 3-202

Caldow C, Monaco ME, Pittman SJ, Kendall MS, Goedeke TL, et al. 2015. Biogeographic assessments: A framework for information synthesis in marine spatial planning. *Marine Policy* 51: 423-32

Callaway R, Engelhard G, Dann J, Cotter J, Rumohr H. 2007. A century of North Sea epibenthos and trawling: comparison between 1902-1912, 1982-1985 and 2000. *Marine Ecology Progress Series* 346: 27-43

Caruana R, Elhawary M, Munson A, Riedewald M, Sorokina D, et al. 2006. *Mining citizen science data to predict orevalence of wild bird species*. Presented at Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining

Carwardine J, Rochester WA, Richardson KS, Williams KJ, Pressey RL, Possingham HP. 2007. Conservation planning with irreplaceability: Does the method matter? *Biodiversity and Conservation* 16: 245-58

Center for International Earth Science Information Network. 2014. Gridded Population of the World, Version 4 (GPWv4). http://www.ciesin.columbia.edu/data/gpw-v4

Chassot E, Bonhommeau S, Dulvy NK, Mélin F, Watson R, et al. 2010. Global marine primary production constrains fisheries catches. *Ecology letters* 13: 495-505

Cheney B, Thompson PM, Ingram SN, Hammond PS, Stevick PT, et al. 2013. Integrating multiple data sources to assess the distribution and abundance of bottlenose dolphins Tursiops truncatus in Scottish waters. *Mammal Review* 43: 71-88

Cheung WW, Lam VW, Sarmiento JL, Kearney K, Watson R, Pauly D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries* 10: 235-51

Cinner JE, Graham NAJ, Huchery C, Macneil MA. 2013. Global Effects of Local Human Population Density and Distance to Markets on the Condition of Coral Reef Fisheries. *Conservation Biology* 27: 453-58

Clarke C, Lea J, Ormond R. 2012. Comparative abundance of reef sharks in the Western Indian Ocean. In *Proceedings of the 12th International Coral Reef Symposium*, pp. 9-13

Cliff AD, Ord JK. 1968. *The problem of spatial autocorrelation*: University

Collen B, Nicholson E. 2014. Taking the measure of change. *Science* 346: 166-67

Collin A, Hench J, Planes S. 2012. A novel spaceborne proxy for mapping coral cover. *12th International Coral Reef Symposium*: 9-13

Costello MJ, Vanden Berghe E. 2006. 'Ocean biodiversity informatics': a new era in marine biology research and management. *Marine Ecology Progress Series* 316: 203-14

Daw TM, Cinner JE, McClanahan TR, Graham NAJ, Wilson SK. 2011. Design factors and socioeconomic variables associated with ecological responses to fishery closures in the Western Indian Ocean. *Coastal Management* 39: 412-24

Deville P, Linard C, Martin S, Gilbert M, Stevens FR, et al. 2014. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* 111: 15888-93

Devlin M, Brodie J, Wenger A, Silva E, Alvarez- JG. 2012. Extreme weather conditions in the Great Barrier Reef : Drivers of change ? In *12th International Coral Reef Symposium*, pp. 9-13

Diniz ‐ Filho JAF, Bini LM, Hawkins BA. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global ecology and Biogeography* 12: 53-64

Dormann C, McPherson J, Araújo M, Bivand R, Bolliger J, et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609-28

Dornelas M, Gotelli NJ, McGill B, Shimadzu H, Moyes F, et al. 2014. Assemblage time series reveal biodiversity change but not systematic loss. *Science* 344: 296-99

Duffy JE, Reynolds PL, Boström C, Coyer JA, Cusson M, et al. 2015. Biodiversity mediates top-down control in eelgrass ecosystems: A global comparative-experimental approach. *Ecology Letters*: in press

Durack PJ, Wijffels SE. 2010. Fifty-Year trends in global ocean salinities and their relationship to broad-scale warming. *Journal of Climate* 23: 4342-62

Edgar GJ, Banks S, Bensted-Smith R, Calvopiña M, Chiriboga A, et al. 2008. Conservation of threatened species in the Galapagos Marine Reserve through identification and protection of marine Key Biodiversity Areas. *Aquatic Conservation: Marine and Freshwater Ecosystems* 18: 955-68

Edgar GJ, Banks SA, Bessudo S, Cortés J, Guzmán HM, et al. 2011. Variation in reef fish and invertebrate communities with level of protection from fishing across the Eastern Tropical Pacific seascape. *Global Ecology and Biogeography* 20: 730-43

Edgar GJ, Barrett NS. 1999. Effects of the declaration of marine reserves on Tasmanian reef fishes, invertebrates and plants. *Journal of Experimental Marine Biology and Ecology* 242: 107-44

Edgar GJ, Barrett NS. 2000. Impact of the Iron Baron oil spill on subtidal reef assemblages in Tasmania. *Marine Pollution Bulletin* 40: 36-49

Edgar GJ, Stuart-Smith RD. 2009. Ecological effects of marine protected areas on rocky reef communities: a continental-scale analysis. *Marine Ecology Progress Series* 388: 51-62

Edgar GJ, Stuart-Smith RD. 2014. Systematic global assessment of reef fish communities by the Reef Life Survey program. *Scientific Data* 1:140007: 1-8

Edgar GJ, Stuart-Smith RD, Willis TJ, Kininmonth S, Banks S, et al. 2014. Global conservation outcomes depend on marine protected areas with five key features. *Nature* 506: 216-20

Eken G, Bennun L, Brooks TM, Darwall W, Fishpool LDC, et al. 2004. Key Biodiversity Areas as site conservation targets. *Bioscience* 54: 1110-18

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802-13

Engelhard GH, Righton DA, Pinnegar JK. 2014. Climate change and fishing: a century of shifting distribution in North Sea cod. *Global Change Biology* 20: 2473-83

Fewster RM, Buckland ST, Siriwardena GM, Baillie SR, Wilson JD. 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology* 81: 1970-84

Folke C. 2015. Globalization , marine regime shifts and the Soviet Union. *Philosophical Transactions of the Royal Society B* 370: 20130278

Fort J, Beaugrand G, Grémillet D, Phillips RA. 2012. Biologging, remotely-sensed oceanography and the continuous plankton recorder reveal the environmental determinants of a seabird wintering hotspot. *PLoS ONE* 7

Francisco-Ramos V, Arias-González JE. 2013. Additive Partitioning of Coral Reef Fish Diversity across Hierarchical Spatial Scales throughout the Caribbean. *PLoS ONE* 8

Fraser HM, Greenstreet S, Piet GJ. 2007. Taking account of catchability in groundfish survey trawls: Implications for estimating demersal fish biomass. *ICES Journal of Marine Science* 64: 1800-19

Fulton E, Smith ADM, Punt AE. 2005. Which ecological indicators can robustly detect effects of fishing? . *Journal of Marine Science* 62: 540-51

Gerritsen H, Lordan C. 2011. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES Journal of Marine Science* 68: 245-52

Graham NAJ, Dulvy NK, Jennings S, Polunin NVC. 2005. Size-spectra as indicators of the effects of fishing on coral reef fish assemblages. *Coral Reefs* 24: 118-24

Graham NAJ, Jennings S, MacNeil A, Mouillot D, Wilson SK. 2015. Predicting climate-driven regime shifts versus rebound potential in coral reefs. *Nature* 518: 94-97

Grassle JF. 2000. The Ocean Biogeographic Information Database (OBIS): an on-line, worldwide atlas for accessing, modeling and mapping marine biological data in a multidimensional geographic context. *Oceanography* 13: 5-7

Group on Earth Observations Biodiversity Observation Network. 2011. *Adequacy of Biodiversity Observation Systems to support the CBD 2020 Targets. A report prepared by the Group on Earth Observations Biodiversity Observation Network (GEO BON) for the Convention on Biological Diversity*. Pretoria, South Africa: Group on Earth Observations Biodiversity Observation Network; http://www.earthobservations.org/documents/cop/bi_geobon/2011_cbd_adequacy_report.pdf

Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, et al. 2012. An index to assess the health and benefits of the global ocean. *Nature* 488: 615-20

Halpern BS, Selkoe KA, Micheli F, Kappel CV. 2007. Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. *Conservation Biology* 21: 1301-15

Halpern BS, Walbridge S, Selkoe Ka, Kappel CV, Micheli F, et al. 2008. A global map of human impact on marine ecosystems. *Science (New York, N.Y.)* 319: 948-52

Hardisty A, Roberts D, Community TBI. 2013. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology* 13: 16

Hendee J, Gramer LJ, Heron SF, Jankulak M, Shoemaker M, et al. 2012. Wireless architectures for coral reef environmental monitoring. In *12th International Coral Reef Symposium*, pp. 9-13

Hochachka WM, Caruana R, Fink D, Munson ART, Riedewald M, et al. 2007. Data ‐ Mining Discovery of Pattern and Process in Ecological Systems. *The Journal of Wildlife Management* 71: 2427-37

Hodgson G. 1999. A global assessment of human effects on coral reefs. *Marine Pollution Bulletin* 38: 345-55

Holt BG, Rioja-Nieto R, Aaron Macneil M, Lupton J, Rahbek C. 2013. Comparing diversity data collected using a protocol designed for volunteers with results from a professional alternative. *Methods in Ecology and Evolution* 4: 383-92

Hosie GW, Fukuchi M, Kawaguchi S. 2003. Development of the Southern Ocean Continuous Plankton Recorder survey. *Progress in Oceanography* 57: 263-83

Hosoda S, Ohira T, Nakamura T. 2008. A monthly mean dataset of global oceanic temperature and salinity derived from Argo float observations. *JAMSTEC Report of Research and Development* 8: 47-59

Illian JB, Sørbye SH, Rue H. 2012. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics* 6: 1499-530

Jackson JBC. 2008. Ecological extinction and evolution in the brave new ocean. *Proceedings of the National Academy of Sciences of the United States of America* 105: 11458-65

Jennings S, Dulvy NK. 2005. Reference points and reference directions for size-based indicators of community structure. *ICES Journal of Marine Science* 62: 397-404

Jiguet F, Devictor V, Julliard R, Couvet D. 2012. French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica* 44: 58-66

Jones JPG, Collen B, Atkinson G, Baxter PWJ, Bubb P, et al. 2011. The why, what, and how of global biodiversity indicators beyond the 2010 target. *Conservation Biology* 25: 450-57

Jones MC, Cheung WWL. 2014. Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES Journal of Marine Science*: DOI: 10.1093/icesjms/fsu172

Keith SA, Webb TJ, Böhning-Gaese K, Connolly SR, Dulvy NK, et al. 2012. What is macroecology? *Biology letters* 8: 904-06

Kerr JT, Kharouba HM, Currie DJ. 2007. The macroecological contribution to global change solutions. *Science* 316: 1581-84

Kéry M, Gardner B, Monnerat C. 2010. Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*: 1851-62

King R. 2012. A review of Bayesian state-space modelling of capture–recapture–recovery data. *Interface Focus* 2: 190-204

Kininmonth S. 2007. Considerations in Establishing Environmental Sensor Networks. In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp. 687-91: Ieee

Kohler KE, Gill SM. 2006. Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. *Computers and Geosciences* 32: 1259-69

Korneliussen RJ, Ona E. 2002. An operational system for processing and visualizing multi-frequency acoustic data. *ICES Journal of Marine Science: Journal du Conseil* 59: 293-313

Kühn I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions* 13: 66-69

Langhammer PF, Bakarr MI, Bennun LA, Brooks TM, Clay RP, et al. 2007. *Identification and gap analysis of Key Biodiversity Areas: Targets for comprehensive protected area systems*. Gland, Switzerland: IUCN Best Practice Protected Areas Guidelines Series No. 15, IUCN. 1-179 pp.

Leaper R, Dunstan P, Foster S, Barrett NS, Edgar GJ. 2012. Comparing large-scale bioregions and fine-scale community-level biodiversity predictions from subtidal rocky reefs across south-eastern Australia. *Journal of Applied Ecology* 49: 851-60

Lee J, South AB, Jennings S. 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES Journal of Marine Science* 67: 1260-71

Levy O, Ball BA, Bond-Lamberty B, Cheruvelil KS, Finley AO, et al. 2014. Approaches to advance scientific understanding of macrosystems ecology. *Frontiers in Ecology and the Environment* 12: 15-23

Leyk S, Buttenfield BP, Nagle NN. 2013. Modeling ambiguity in census microdata allocations to improve demographic small area estimates. *Transactions in GIS* 17: 406-25

Link JS, Yemane D, Shannon LJ, Coll M, Shin Y-J, et al. 2010. Relating marine ecosystem indicators to fishing and environmental drivers: an elucidation of contrasting responses. *ICES Journal of Marine Science: Journal du Conseil* 67: 787-95

Liu G, Eakin CM, Rauenzahn JL, Christensen TRL, Scott F, et al. 2012. NOAA Coral Reef Watch ' s Decision Support System for Coral Reef Management. In *12th International Coral Reef Symposium*, pp. 9-13

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Andrew Royle J, Langtimm CA. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83: 2248-55

Mallet D, Pelletier D. 2014. Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952-2012). *Fisheries Research* 154: 44-62

Mallet D, Wantiez L, Lemouellic S, Vigliola L, Pelletier D. 2014. Complementarity of Rotating Video and Underwater Visual Census for Assessing Species Richness, Frequency and Density of Reef Fish on Coral Reef Slopes. *PLoS ONE* 9: e84344

Marin-Perianu M, Chatterjea S, Marin-Perianu R, Bosch S, Dulman S, et al. 2008. Wave monitoring with wireless sensor networks. *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*: 611-16

Matabos M, Bui AOV, Mihály S, Aguzzi J, Juniper SK, Ajayamohan RS. 2014. High-frequency study of epibenthic megafaunal community dynamics in Barkley Canyon: A multi-disciplinary approach using the NEPTUNE Canada network. *Journal of Marine Systems* 130: 56-68

McCauley DJ, Pinsky ML, Palumbi SR, Estes JA, Joyce FH, Warner RR. 2015. Marine defaunation: Animal loss in the global ocean. *Science* 347: 1255641

McKenzie L, Collier C, Waycott M, Unsworth R, Yoshida R, Smith N. 2012. Monitoring inshore seagrasses of the GBR and responses to water quality. *12th International Coral Reef Symposium*: 9-13

Meinshausen N. 2006. Quantile regression forests. *The Journal of Machine Learning Research* 7: 983-99

Mellin C, Delean S, Caley MJ, Edgar GJ, Meekan MG, et al. 2011. Effectiveness of biological surrogates for predicting patterns of marine biodiversity. *PLoS One* 6: e20141. doi:10.1371/journal.pone.0020141

Michener WK, Jones MB. 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in ecology & evolution* 27: 85-93

Miloslavich P, Webb TJ, Snelgrove P, Vanden Berghe E, Kaschner K, et al. in press. Extent of assessment of marine biological diversity. In *World Ocean Assessment*

Monk J. 2014. How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? *Fish and Fisheries* 15: 352-58

Mora C, Aburto-Oropeza O, Ayotte PM, Banks S, Bauman AG, et al. 2011. Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. *PLoS Biology* 9: e1000606

Nicholson E, Collen B, Barausse A, Blanchard JL, Costelloe BT, et al. 2012. Making robust policy decisions using global biodiversity indicators. *PLoS ONE* 7: e41128

Nicholson MD, Jennings S. 2004. Testing candidate indicators to support ecosystem-based management: the power of monitoring surveys to detect temporal trends in fish community metrics. *ICES Journal of Marine Science: Journal du Conseil* 61: 35-42

O'Connell AF, Talancy NW, Bailey LL, Sauer JR, Cook R, Gilbert AT. 2006. Estimating site occupancy and detection probability parameters for meso-and large mammals in a coastal ecosystem. *Journal of Wildlife Management* 70: 1625-33

Patil AP, Okiro EA, Gething PW, Guerra CA, Sharma SK, et al. 2009. Defining the relationship between Plasmodium falciparum parasite rate and clinical disease: statistical models for disease burden estimation. *Malar J* 8: 186

Pauly D, Christensen V, Walters C. 2000. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. *ICES Journal of Marine Science [ICES J. Mar. Sci.].* 57: 697-706

Pauly D, Watson R, Alder J. 2005. Global trends in world fisheries: impacts on marine ecosystems and food security. *Philosophical Transactions Of The Royal Society B-Biological Sciences* 360: 5-12

Pereira HM, Ferrier S, Walters M, Geller GN, Jongman RHG, et al. 2013. Essential Biodiversity Variables. *Science* 339: 277-78

Pesaresi M, Huadong G, Blaes X, Ehrlich D, Ferri S, et al. 2013. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6: 2102-31

Porter JH, Hanson PC, Lin C-C. 2012. Staying afloat in the sensor data deluge. *Trends in ecology & evolution* 27: 121-29

Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson Ka. 2007. Conservation planning in a changing world. *Trends in ecology & evolution* 22: 583-92

Raiter KG, Possingham HP, Prober SM, Hobbs RJ. 2014. Under the radar: mitigating enigmatic ecological impacts. *Trends in ecology & evolution* 29: 635-44

Rasmussen CE. 2006. Gaussian processes for machine learning.

Raudenbush SW, Yang M-L, Yosef M. 2000. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of computational and Graphical Statistics* 9: 141-57

Ready J, Kaschner K, South AB, Eastwood PD, Rees T, et al. 2010. Predicting the distributions of marine organisms at the global scale. *Ecological Modelling* 221: 467-78

Resilience Alliance and Santa Fe Institute. 2004. Thresholds and alternate states in ecological and social-ecological systems. *http://www.resalliance.org/index.php/thresholds_database*

Ricard D, Minto C, Jensen OP, Baum JK. 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish And Fisheries* 13: 380-98

Rice JC. 2000. Evaluating fishery impacts using metrics of community structure. *ICES Journal of Marine Science* 57: 682-88

Richards Z. 2014. The status of hard coral communities at Kosrae, Micronesia. *Marine Biodiversity*: 1-12

Richardson AJ, Poloczanska ES. 2008. Ocean science: under-resourced, under threat. *Science* 320: 1294-95

Richardson AJ, Walne AW, John AWG, Jonas TD, Lindley JA, et al. 2006. Using continuous plankton recorder data. *Progress In Oceanography* 68: 27-74

Robertson DR. 2008. Global biogeographical data bases on marine fishes: caveat emptor. *Diversity and Distributions* 14: 891-92

Rodrigues ASL, Brooks TM. 2007. Shortcuts for biodiversity conservation planning: the effectiveness of surrogates. . *Annual Review of Ecology, Evolution, and Systematics* 38: 713-37

Roemmich D, Gould WJ, Gilson J. 2012. 135 years of global ocean warming between the Challenger expedition and the Argo Programme. *Nature Climate Change* 2: 425-28

Roemmich D, Johnson GC, Riser S, Davis R, Gilson J, et al. 2009. The Argo Program: Observing the global ocean with profiling floats. *Oceanography* 22: 34-43

Rogers-Bennett L, Haaker PL, Karpov KA, Kushner DJ. 2002. Using Spatially Explicit Data to Evaluate Marine Protected Areas for Abalone in Southern California. *Conservation Biology [Conserv. Biol.].* 16: 1308-17

Royle JA, Kéry M, Gautier R, Schmid H. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs* 77: 465-81

Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 71: 319-92

Ruttenberg BI, Schofield PJ, Akins JL, Acosta A, Feeley MW, et al. 2012. Rapid invasion of Indo-Pacific lionfishes (Pterois volitans and Pterois miles) in the Florida Keys, USA: evidence from multiple pre- and post-invasion data sets. *Bulletin of Marine Science* 88: 1051-59

Shamir L, Yerby C, Simpson R, von Benda-Beckmann AM, Tyack P, et al. 2014. Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *The Journal of the Acoustical Society of America* 135: 953-62

Shin Y-J, Rochet M-J, Jennings S, Field JG, Gislason H. 2005. Using size-based indicators to evaluate the ecosystem effects of fishing. *ICES Journal of Marine Science: Journal du Conseil* 62: 384-96

Shotton J, Sharp T, Kohli P, Nowozin S, Winn J, Criminisi A. 2013. *Decision jungles: Compact and rich models for classification*. Presented at Advances in Neural Information Processing Systems

Simpson SD, Jennings S, Johnson MP, Blanchard JL, Schön P-J, et al. 2011. Continental Shelf-Wide Response of a Fish Assemblage to Rapid Warming of the Sea. *Current Biology* 21: 1565-70

Smale DA, Langlois TJ, Kendrick GA, Meeuwig JJ, Harvey ES. 2011. From fronds to fish: the use of indicators for ecological monitoring in marine benthic ecosystems, with case studies from temperate Western Australia. *Reviews in Fish Biology and Fisheries* 21: 311-37

Song C, Qu Z, Blumm N, Barabási A-L. 2010. Limits of predictability in human mobility. *Science (New York, N.Y.)* 327: 1018-21

Spear LB, Ainley DG, Hardesty BD, Howell SN, Webb SW. 2004. Reducing biases affecting at-sea surveys of seabirds: use of multiple observer teams. *Marine Ornithology* 32: 147-57

Stattersfield AJ, Crosby MJ, Long AJ, Wege DC. 1998. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. Cambridge: BirdLife International

Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *Plos One* 10: e0107042

Stuart-Smith RD, Bates AE, Lefcheck JS, Duffy JE, Baker SC, et al. 2013. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* 501: 539-42

Sutherland WJ, Adams WM, Aronson RB, Aveling R, Blackburn TM, et al. 2009. One hundred questions of importance to the conservation of global biological diversity. *Conservation Biology* 23: 557-67

Sweatman H, Delean S, Syms C. 2011. Assessing loss of coral cover on Australia's Great Barrier Reef over two decades, with implications for longer-term trends. *Coral Reefs* 30: 521-31

Thomson SCW. 1880. Report on the Scientific Results of the Voyage of H.M.S. Challenger During the Years 1872-76. 525

Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, et al. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature* 466: 1098-101

Treml E, Fidelman P, Kininmonth S, Ekstrom J, Bodin Ö. 2015. Analysing the (Mis)Fit between Institutional and ecological networks of the Coral Triangle. *Global Environmental Change* 31: 263-71

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F, De Clerck O. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modeling. *Global Ecology and Biogeography* 21: 272-81

Tyler EH, Somerfield PJ, Berghe EV, Bremner J, Jackson E, et al. 2012. Extensive gaps and biases in our knowledge of a well‐known fauna: implications for integrating biological traits into macroecology. *Global Ecology and Biogeography* 21: 922-34

Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP. 2003. Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications* 13: 1790-801

Vandepitte L, Bosch S, Tyberghein L, Waumans F, Vanhoorne B, et al. 2015. Fishing for data and sorting the catch: assessing the data quality, completeness and fitness for use of data in marine biogeographic databases. *Database* 2015: bau125

Varjopuro R, Andrulewicz E, Blenckner T, Dolch T, Heiskanen A-s, Pihlajamäki M. 2014. Coping with persistent environmental problems : systemic delays in reducing eutrophication of the Baltic Sea. *Ecology & Society* 19: 48

Walters CJ, Holling CS. 1990. Large-scale management experiments and learning by doing. *Ecology* 71: 2060-68

Webb TJ. 2012. Marine and terrestrial ecology: unifying concepts, revealing differences. *Trends in ecology & evolution* 27: 535-41

Webb TJ, Tyler EHM, Somerfield PJ. 2009. Life history mediates large-scale population ecology in marine benthic taxa. *Marine Ecology Progress Series* 396: 293-306

Webb TJ, Vanden Berghe E, O'Dor R. 2010. Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PLoS ONE* 5: e10223

Wikle CK. 2003. Hierarchical Models in Environmental Science. *International Statistical Review* 71: 181-99

Wilkinson B, Allen M. 1999. *Parallel programming*: Prentice hall New Jersey

Wilkinson C. 2008. Status of Coral Reefs of the World. 296

Williams R, Hedley SL, Hammond PS. 2006. Modeling distribution and abundance of Antarctic baleen whales using ships of opportunity. *Ecology and Society* 11

Young IR, Zieger S, Babanin aV. 2011. Global trends in wind speed and wave height. *Science (New York, N.Y.)* 332: 451-55

Zhao H, Ai S, Lv Z, Li B. 2010. Parallel accessing massive NetCDF data based on mapreduce. In *Web Information Systems and Mining*, pp. 425-31: Springer

Zieger S, Stieglitz T, Kininmonth S. 2009. Mapping reef features from multibeam sonar data using multiscale morphometric analysis. *Marine Geology* 264: 209-17

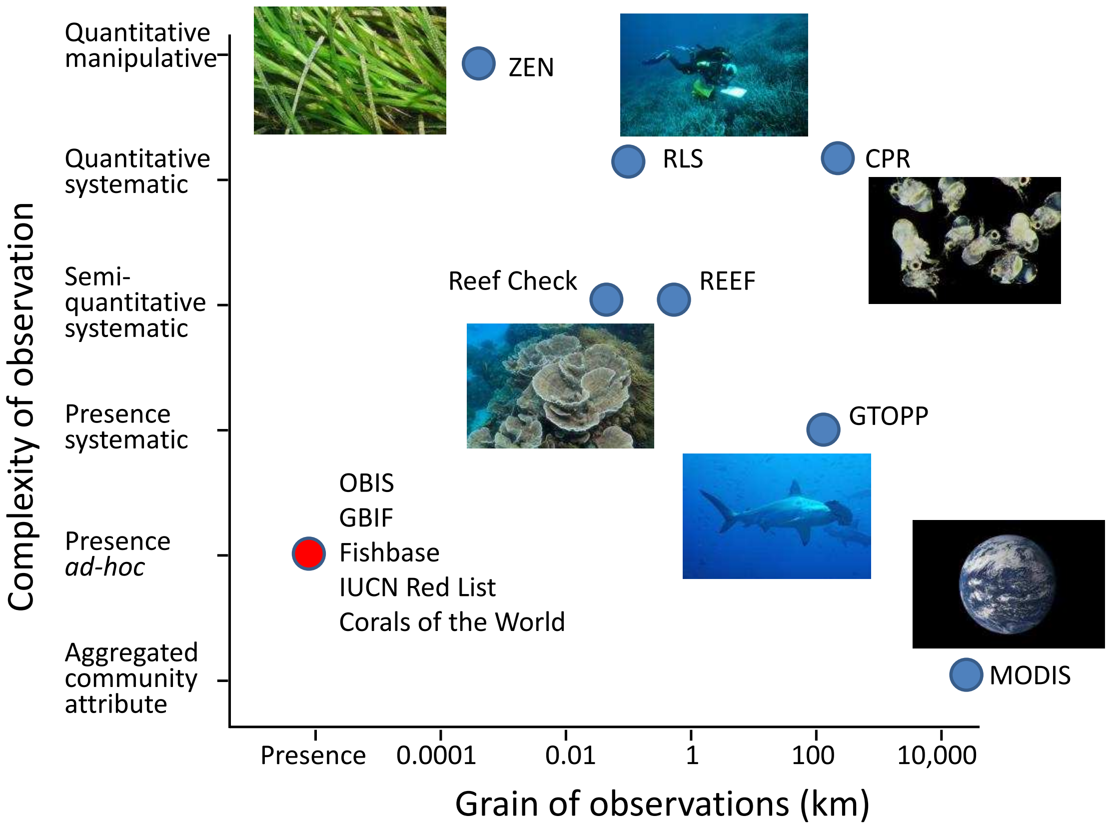

Fig. 1. Relationship between span of replicate observations ('grain') and biological detail obtained ('complexity') for major global marine observing systems discussed in this review. High complexity is only achieved by systematic observation systems with small grain, such as the Zostera Experimental Network (ZEN), where plots of 50 cm diameter are manipulated. By contrast, MODIS coverage of ocean colour, a proxy for phytoplankton biomass, fully encompasses the globe. RLS: Reef Life Survey; CPR: Continuous Plankton Recorder; REEF: Reef Environmental Education Foundation; GTOPP: Global Tagging of Pelagic Predators; OBIS: Ocean Biogeographic Information System; GBIF: Global Biodiversity Information Facility.

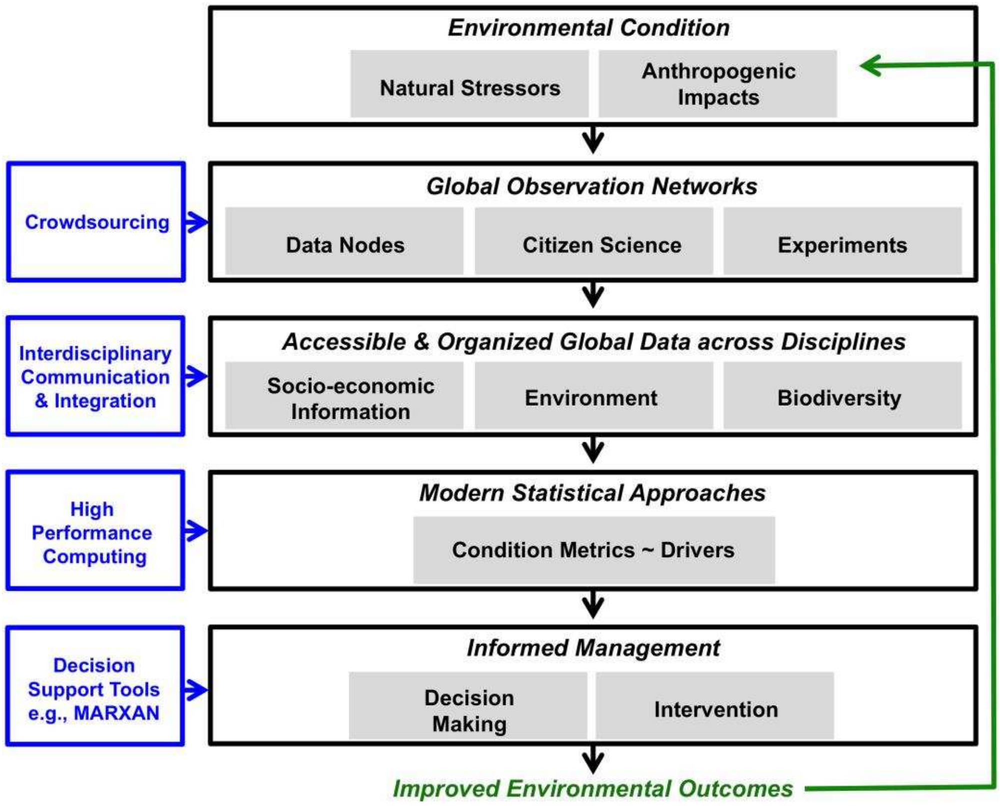

Fig. 2. Flow chart describing links between environmental condition and the observation, analysis and management elements that constitute this review.

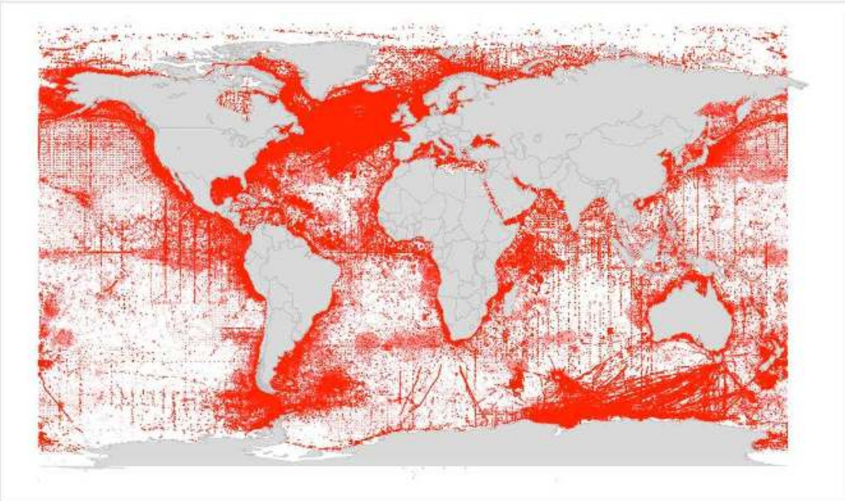

Fig. 3. Distribution of locations with OBIS species' records.

Comment [GE2]: Note to editor: this figure can be deleted if paper overlength

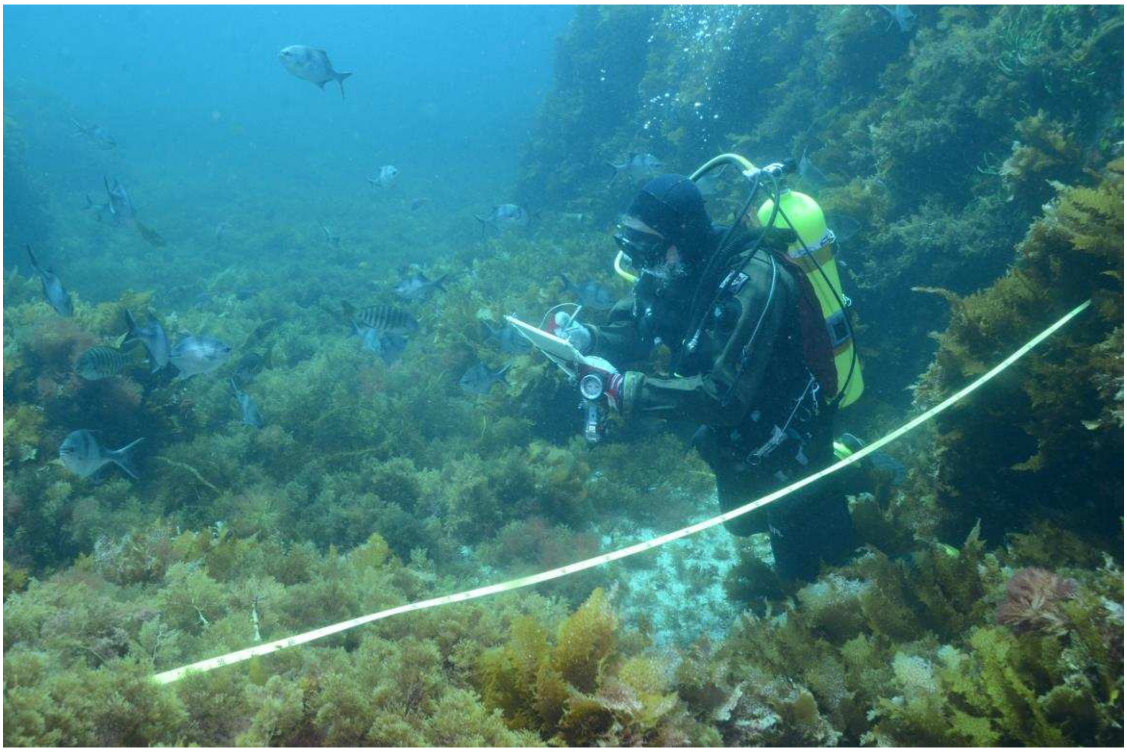

Fig. 4. Diving citizen scientist undertaking Reef Life Survey fish count.

Comment [GE3]: Note to editor: this figure can be deleted if paper overlength

Table 1. Useful analytical and data management approaches for large marine ecological datasets

Statistical Method	Application Description	Example
Hierarchical models	**Accounting for conditional dependence in the variance of the data**	
Mixed effects	Uneven variability is partitioned using metadata	(Bird et al 2014)
GAM(M)	Response data are modelled as a function of smoothed prediction data	(Fewster et al 2000)
Occupancy detection	Replicate observations within sampling units allows estimation of probability of detecting a species given its presence	(O'Connell et al 2006)
Capture-recapture	Replicate attempts to capture tagged animals yields an estimate of capture probability	(Cheney et al 2013)
Multiple observer	Estimate detection error by observing how well two independent observations of occurrence overlap	(Spear et al 2004)
Line-transect methods	Infer density based on the assumption that probability of detection falls off with distance	(Buckland et al 1993, Burnham et al 1980)
Integrated population models	Measuring the same population using multiple approaches allows more accurate inference on population size	(Besbeas et al 2002)
Spatial methods	**Accounting for the spatial non-independence of data points**	(Reviewed in Dormann et al 2007)
Geographically-weighted regression	Performs linear regression at local scales to quantify relationships that vary in space	(Brunsdon et al 1998)
Kriging	Predicted values between observed data points are interpolated using a gaussian process with pre-set parameters	(Jiguet et al 2012)

Bayesian Geostatistical models	Formulating the spatial dependence of data points in a Bayesian framework allows exploration of complex hierarchical dependencies	Chang and Yuan, 2014
Integrated Nested Laplace approximation	INLA allows Bayesian hierarchical models to converge more quickly by avoiding MCMC	(Illian et al 2012, Rue et al 2009)
Environmental envelope	Observed species ranges are related to the range of environmental conditions experienced throughout their range.	(Cheung et al 2009)
Principle Components Analysis	PCA allows detection and quantification of spatial patterns over different scales	(Borcard & Legendre 2002)
Machine learning	**Exploration of how response data are explained by many predictors**	
Random forests	randomly allocated classification rules show how combinations of covariates predict response data	(Stuart-Smith et al 2013)
Quantile regression forests	the distribution of branching algorithms chosen in random forests provides estimates of uncertainty in their predictions	(Meinshausen 2006)
Random Jungles	probabilistic clustering of branching rules allows efficient exploration of large sets of predictor variables	(Shotton et al 2013)
Boosted regression trees	Classification algorithms performed on predictions from sequential trees allows for more robust predictions	(Hochachka et al 2007)
Gaussian Process model	Smoothed relationships between response data and predictors are modeled as multinomial gaussian distribution in multidimensional space	(Patil et al 2009)
Data management		

HPC	Splitting large computational tasks between multiple computers allows analysis and prediction in large datasets	(Caruana et al 2006)
Distributed analyses	Analytical or processing tasks can often be divided between computers before being aggregated for a final analysis	(Anderson et al 2002)
Crowdsourcing	Pattern recognition tasks can farmed out to volunteers over the internet	(Shamir et al 2014)
Hadoop	The Map-reduce computing framework allows large data files to be logically processed across a distributed network of compute nodes	(Zhao et al 2010)