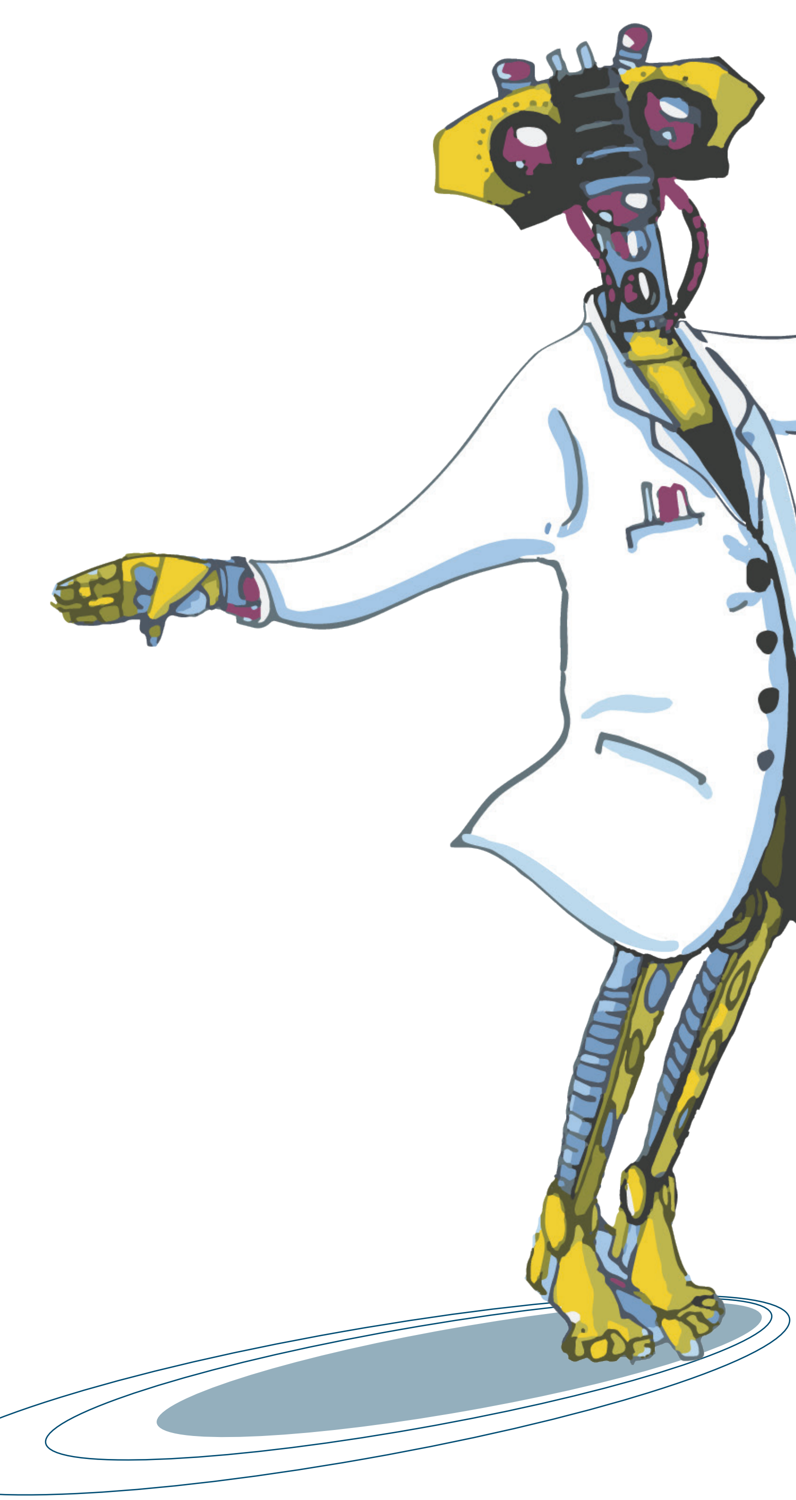
WG149



SCOR WG149
Handbook to support the SCOR

Best Practice Guide for Multiple Drivers Marine Research

March 2019

www.meddle-scor149.org

ISBN 978-1-925646-72-6

DOI 10.25959/5c92fdf0d3c7a

Publication date: 2019

Publisher: University of Tasmania, on behalf of Scientific Committee on Oceanic Research (SCOR)

Authors SCOR WG149 Philip Boyd (Australia), Aurea Ciotti (Brazil), Sinead Collins (UK), Kunshan Gao (China-Beijing), Jean-Pierre Gattuso (France), Marion Gehlen (France), David Hutchins (USA), Christina McGraw (Australia), Jorge Navarro (Chile), and Ulf Riebesell (Germany), Haimanti Biswas (India), Sam Dupont (Sweden), Katharina Fabricius (Australia), Jonathan Havenhand (Sweden), Catriona Hurd (Australia), Haruko Kurihara (Japan), Gorann Nilsson (Norway), Uta Passow (USA), Hans-Otto Pörtner (Germany), and Marcello Vichi (Italy)

Please cite as:
Boyd, PW, Collins, S, Dupont, S, Fabricius, K, Gattuso, J-P, Havenhand, J, Hutchins, DA, McGraw, CM, Riebesell, U, Vichi, M, Biswas, H, Ciotti, A, Dillingham, P, Gao, K, Gehlen, M, Hurd, CL, Kurihawa, H, , Navarro, J, Nilsson, GE, Passow, U and Portner, H-O (2019). SCOR WG149 Handbook to support the SCOR Best Practice Guide for 'Multiple Drivers' Marine Research. http://dx.doi.org/10.25959/5c92fdf0d3c7a

Thank you to Alex Naughton for the "Bot" illustrations (front and back covers, and inside pages)
Design and layout by Louise Bell Graphic Design louise.bell@lightbox42.com.au
Website designed and developed by Ionata Digital www.ionata.com.au

Contents

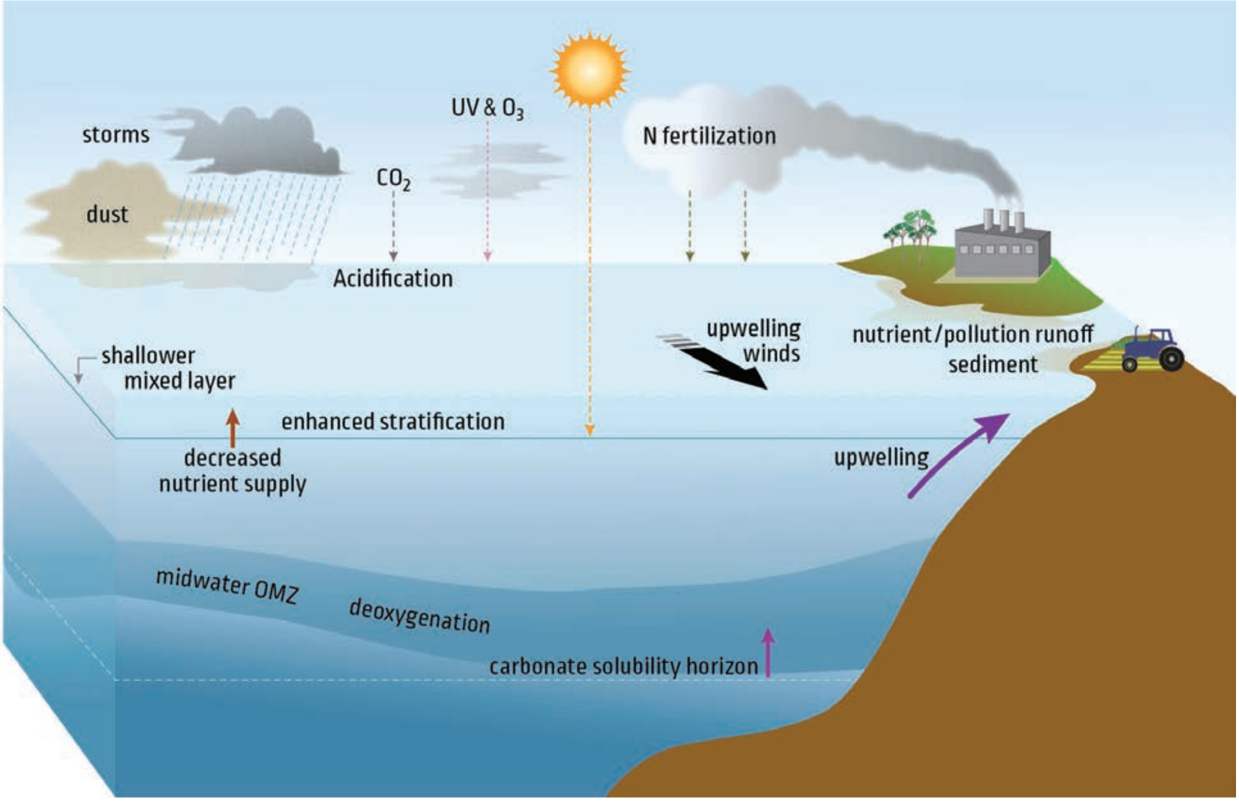
1. INTRODUCTION

Marine species and ecosystems are exposed to a wide range of environmental change – both detrimental (threats) and beneficial – due to human activities. Some of the changes are global, whereas others are regional or local. It is important to distinguish the scale of each threat as the solutions will differ. For example, the mitigation of a global problem requires a global response, which is more difficult to achieve than addressing a local problem with a local response. These wide-ranging changes are often referred to drivers or stressors.

1.1 WHAT ARE MULTIPLE DRIVERS?

The term multiple drivers refers to the concurrent alteration of multiple environmental properties, that are each biologically-influential, by anthropogenic pressures including climate change. These multiple environmental properties are commonly referred to as drivers or stressors, and include temperature, carbon dioxide, pH, oxygen, salinity, density, irradiance and nutrients, eutrophication, UV exposure, and point source pollutants (Figure 1).

The multiple drivers framework represents a complex matrix of changing ocean properties, that will vary from locale to locale, and may also alter with season.



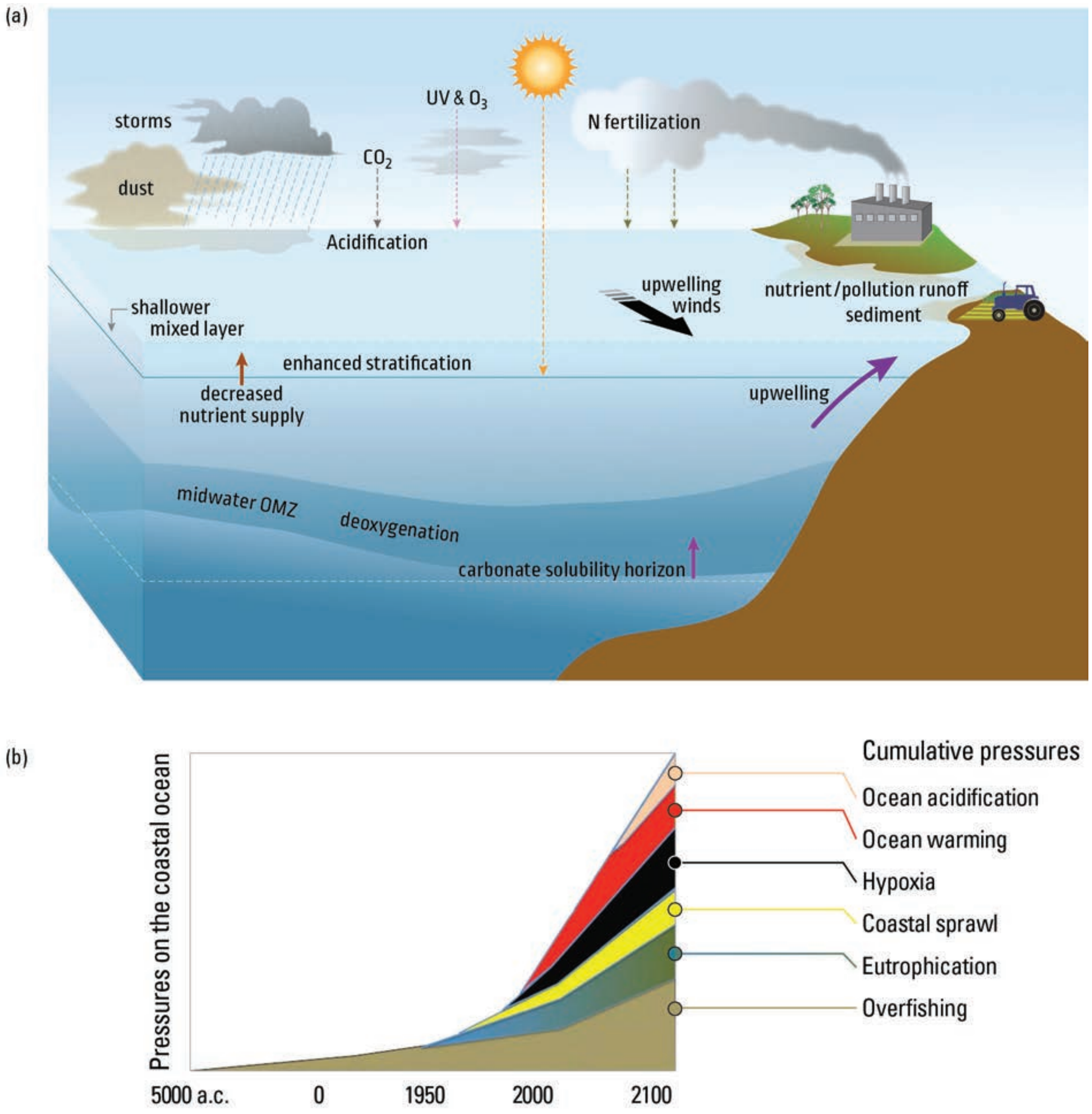


FIGURE 1. Examples of global, regional and local environmental drivers. a) Global drivers are primarily mediated by anthropogenic pressures and include oxygen (deoxygenation), carbon dioxide (acidification) and seawater density (altered stratification). Region drivers include UV radiation (the Ozone hole) and nutrients (atmospherically-transported pollutants). Local drivers include pollutants (for example from point sources) and nutrients or freshwater (terrestrial run-off). From Figure 3a in Boyd *et al*. (2018). b) presents hypothetical time lines for the emergence of cumulative pressures in the coastal zone. Modified from Duarte (2013, In: The Conversation, http://theconversation.com/auditing-the-seven-plagues-of-coastal-ecosystems-13637).

1.2 Translating changing marine conditions into biological outcomes

How do such multiple drivers translate into outcomes for marine life? At the organismal level, environmental drivers such as irradiance, nutrients and carbon dioxide are essential for processes such as photosynthetic carbon fixation or the synthesis of macromolecules. Temperature also plays a key role in setting the rate of most cellular processes. Every species has a certain tolerance to individual drivers, and may be influenced by a different suite of drivers. This can be explained by different factors:

- Adaptation to local conditions (e.g. Vargas *et al*. 2017)
- Life stages
- Mode of nutrition – primary producers versus herbivores

The translation of environmental forcing into biological outcomes can be represented graphically for many of these processes, for example a thermal performance curve summarises how a measure of fitness such as growth rate changes with temperature (Figure 2a). Other physiological metrics include nutrient affinity curves (Figure 2b) or photosynthesis versus irradiance curves. Other examples of these curves are provided on the website under “Learning materials”.

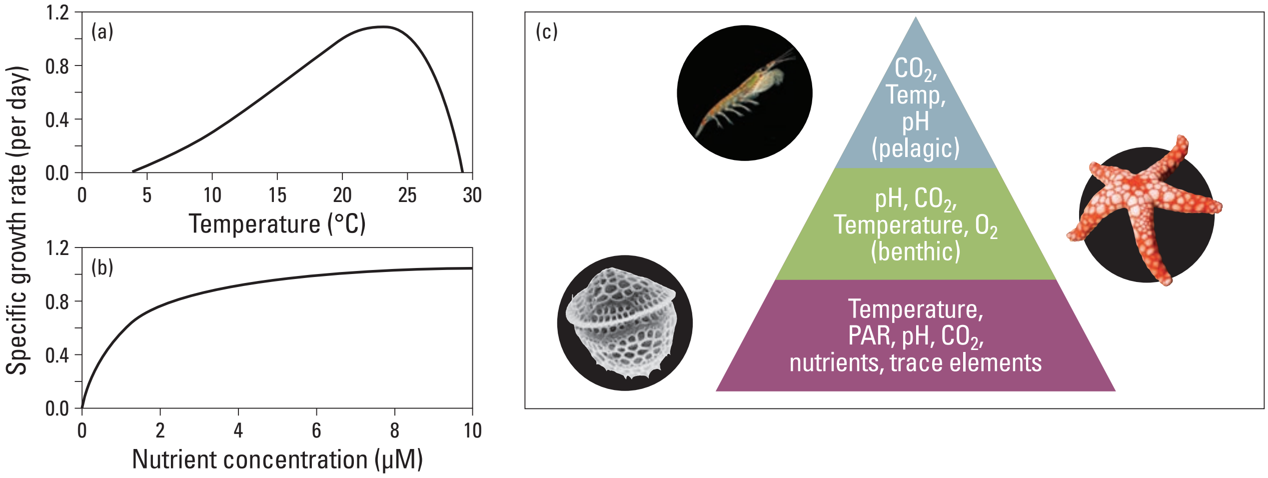


FIGURE 2. Examples of how organism respond (i.e., mode of action) to environmental properties for a) specific growth rate versus temperature; and b) specific growth rate versus nutrient concentration (redrawn from Thomas *et al*., 2017). Panel c) illustrates the point that different species may be influenced by a distinct suite of drivers, that may be linked to their mode of nutrition or their habitat. Krill image courtesy of Australian Antarctic Division.

Performance curves such as those in Figure 2 represent the response of an individual species to a single driver (often following acclimation). These curves can be mapped onto figure 3 which is a ‘cube’ comprising space, time and drivers. This cube from Riebesell and Gattuso (2015) reveals the complex interplay between multiple drivers and physiological responses at the species level, and the need to eventually relate them to ecological (species to ecosystems) and/or evolutionary (acclimated to adapted) scales.

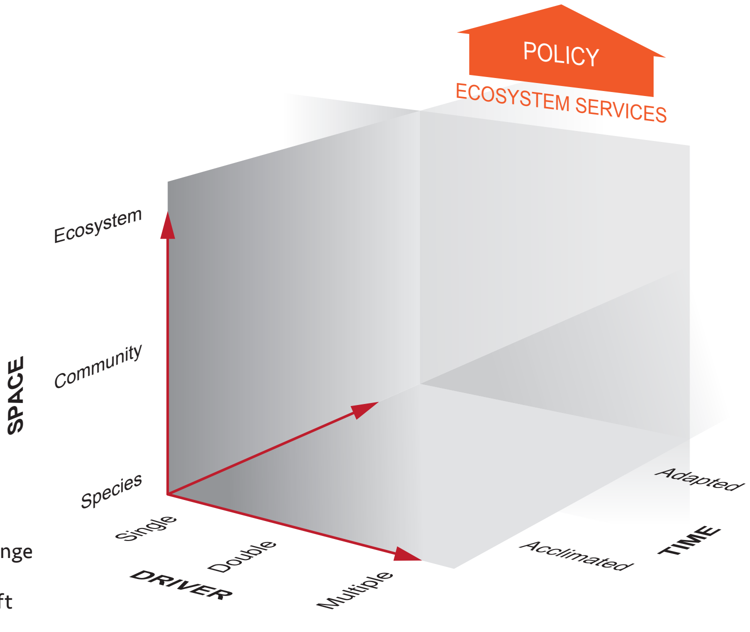


FIGURE 3. Present state of knowledge and knowledge needs: most information on the impacts of ocean change currently available is on acclimated single species/strains under the influence of a single driver (lower left corner). Redrawn from Riebesell and Gattuso (2015).

Examination of each of the axes in Figure 3, in turn, provides an appreciation of the complexity of the challenges. For example, for a single species the individual performance curve for a single driver is influenced (i.e., modulated) by other drivers – for example nutrients. This modulation will depend upon on the mode of action. From a physiological or metabolic standpoint, the challenge is to understand the mechanism(s) of response by the species to understand the interactions between these drivers. In the case of temperature and nutrient supply, at low (i.e., limiting) nutrient concentrations the shape of the thermal performance curve for coastal phytoplankton was altered (for more details see Thomas *et al.* 2017).

Next, the examination of another axis of Figure 3 – that of space (species to community to ecosystem) brings additional challenges. For example, a comparison of the responses, to individual drivers, by organisms occupying different trophic levels, such as from primary producers and herbivores indicates that the relationship with multiple drivers and their influence on physiological performance differs. These differences may occur in two generic ways. First, the suite of drivers influencing organismal physiology may alter, for example most grazers do not photosynthesise (see Figure 2c), so the influence of irradiance on grazers will be different than for primary producers. Another example comes from nutrition – primary producers require dissolved nutrients such as nitrate, whereas grazers consume prey. Second, each trophic level may have different physiological relationships with a common driver. For example, the thermal performance curves of a grazer and a primary producer (for respiration) may differ, resulting in different sensitivities to a warming ocean.

As we consider the relationship between trophic levels in the foodweb, further complexities arise. The physiology of a primary producer is directly influenced by multiple drivers (often, irradiance, nutrients, carbon dioxide and temperature). In contrast the physiology of a herbivore is influenced directly by multiple drivers (temperature, oxygen, carbon dioxide), as well as indirectly by prey quantity and quality (which are set by the physiology of the primary producer). A major challenge is to predict the cumulative outcome of these physiological responses (both direct and indirect) on the functioning and structure of the ecosystem.

There are many other potentially confounding issues on linking the Driver and Space axes in Figure 3, including intra- and inter-species diversity and how it defines ecosystem structure. For further reading see section 7 "Bridging between physiological responses and ecosystem impacts" in Boyd *et al.* (2018). For an introduction into the third axis – Time (i.e., from non-acclimated to acclimated to adapted) in Figure 3 see Section 8 "Evolution under multiple drivers" in Boyd *et al.* (2018).

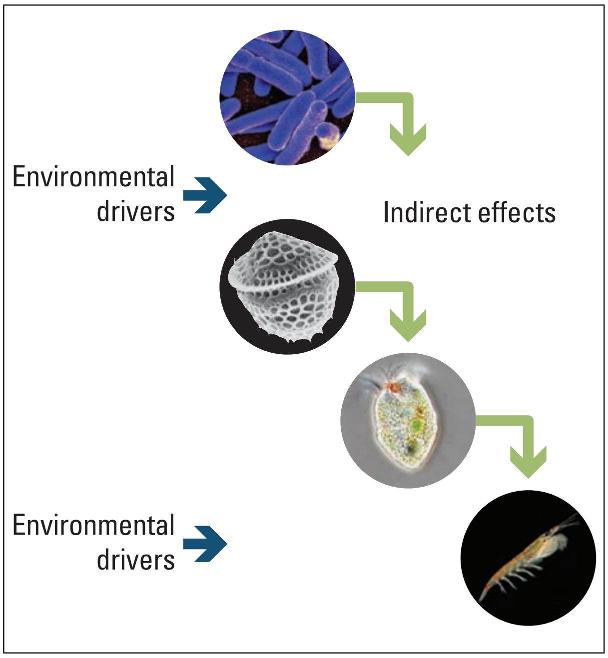


FIGURE 4. A schematic that illustrates how environmental drivers influence organisms occupying different trophic levels. Primary producers such as phytoplankton (upper images) are influenced directly by drivers, whereas grazers are influenced both directly by drivers, and indirectly (via the direct influence of drivers on their prey). In turn, grazers exert a top-down influence on their prey. Krill image courtesy of Australian Antarctic Division.

1.3 PROJECTING BIOLOGICAL RESPONSE(S) TO FUTURE OCEAN CONDITIONS USING PERTURBATION EXPERIMENTS

The wide range of approaches that are available for projecting the response of biota to changing marine conditions are summarised in Figure 5. Here, in this Handbook that accompanies the www-based best Practice Guide, the focus is on using perturbation experiments to probe the response of marine life to future ocean conditions.

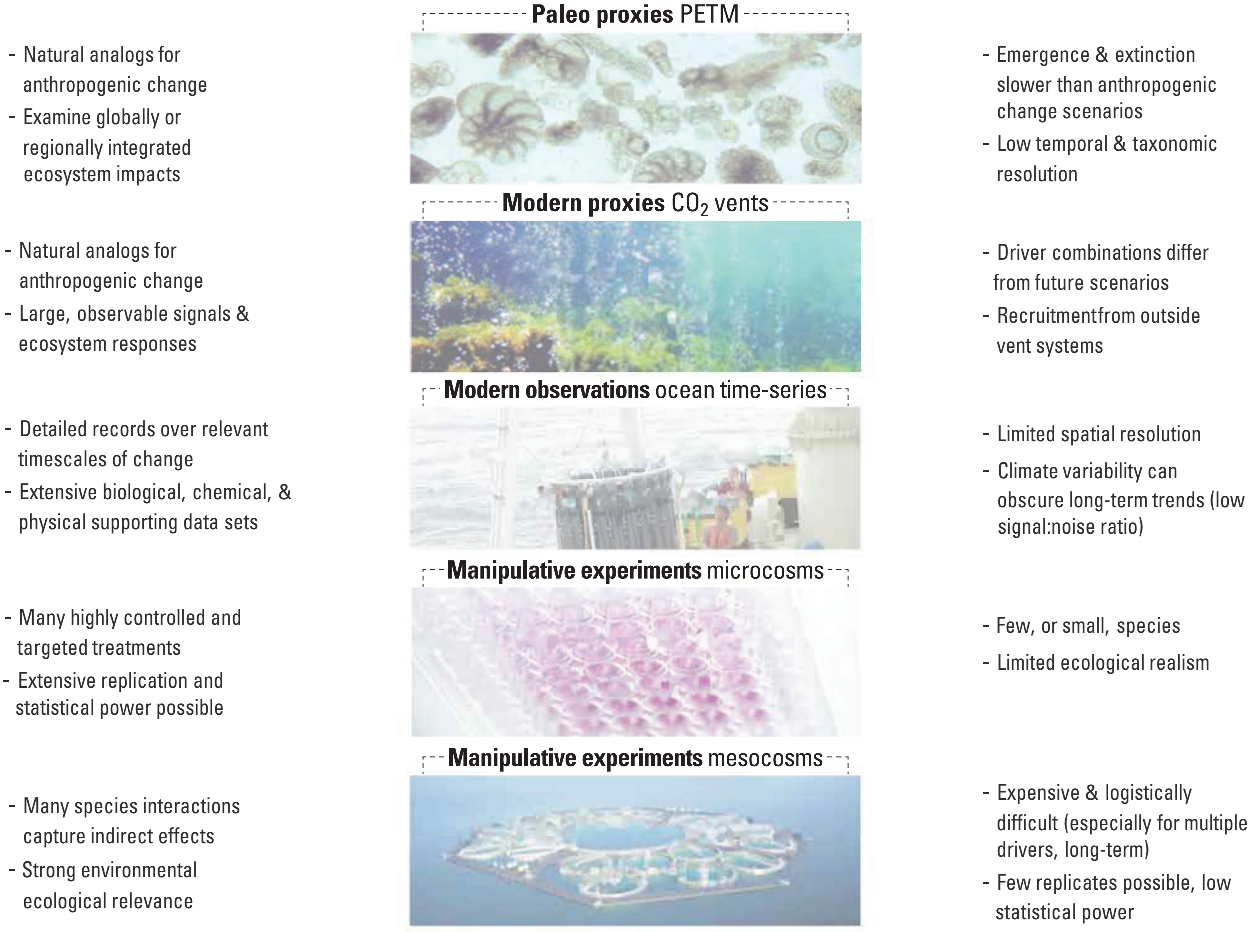


FIGURE 5. Strengths (left column) and limitations (right column) of the five main approaches (centre, rectangles) used to understand the effect of environmental drivers on marine biota. Major approaches include: Paleoceanographic studies of past natural climate shifts (Paleo-Proxies) such as the PETM event ~56 million years ago; Modern natural environments that can serve as proxies of particular anthropogenic change processes (Modern Proxies), such as acidification resulting from seafloor CO_2 vents or regions where naturally low-pH seawater is upwelled; Modern observations that capture extended temporal or spatial aspects of global change, including decadal-scale ocean monitoring sites such as the Bermuda Atlantic Time-Series; Manipulative microcosm experiments often used to carry out controlled experimentation on single species or small communities; and large-volume mesoscosm experiment enclosures and free ocean CO_2 enrichment (FOCE) experiments that are used to manipulate entire marine communities (Figure from Boyd *et al.*, 2018).

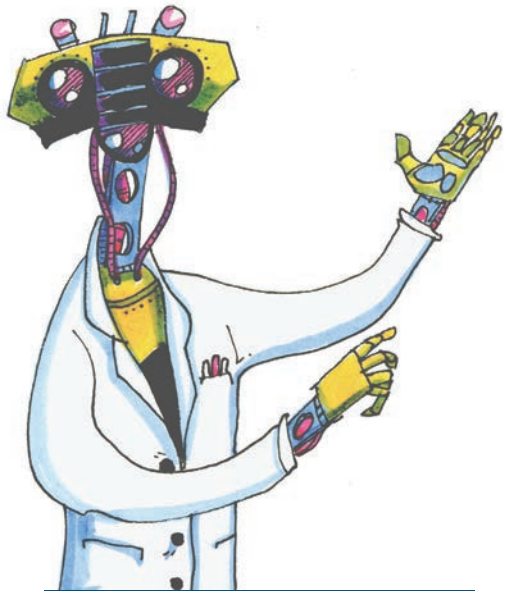
A wide range of perturbation experiments can be used to better understand how multiple drivers influence marine life. The choice of the experimental approach will depend on the question(s) that are being addressed, the study subject, and the locale of the experiment. There are a wide number of permutations that fall within the three axes of the 'cube' in Figure 3, ranging from a single driver, single species acclimated study (such as the thermal performance curve in Figure 2a) to more complex approaches with more than one driver, several trophic levels, and/or many generations (i.e., adaptation studies).

In many cases, it is important to emphasise that some simpler single driver experiments should be conducted first to provide background information that will help to design a better multiple driver experiment. Single stressor studies, for example using a performance curve (Figure 2) provide mechanistic understanding to help identify mode of actions. Such pilot studies will help to assess and plan for the logistical challenges of running more complex (multi-driver) experiments. Preliminary studies also likely result in clearer interpretation of the observed effects (e.g. interactions between drivers that are non-linear) in more complex experiments. Multiple stressors studies are valuable tools to test hypotheses on interactions and modulations.

Table 1 (see next page) provides examples of a wide range of experimental approaches that have been used to investigate the many facets of multiple driver experiments.

It is useful to repeat again:

The choice of the experimental approach will depend on the question(s) that are being addressed, the study subject, and the locale of the experiment.



Experimental Approach	Examples	Benefits	Disadvantages	Main uses
Single driver/mechanistic	Warming (Eppley, 1972)	Intrinsic physiological status; Ability to build models (mathematical or conceptual) from studies of single driver and modes of action, and to iterate this 'loop' (Baretta-Bekker,et al. 1994).	No information on relative influence of other drivers	Reaction norm and Reciprocal interface with models
Single driver/constant conditions	Acidification (Dupont et al., 2008)	Specific response to projected future conditions which can be invaluable if a sole driver is dominant (temperature/coral bleaching, Hughes et al. 2017).	No information on relative influence of other drivers, no information on ecological relevance (lack of realism)	Response to IPCC projections
Single driver/ fluctuations	Acidification (Cornwall et al., 2014, Eriander et al., 2015)	Specific response to projected future conditions and to the influence of natural environmental variability	No information on relative influence of other drivers, no information on ecological relevance (lack of realism)	Response to IPCC projections
Single driver/competition experiment	Acidification (Krause et al., 2012)	Competition as opposed to single species	No information on relative influence of other drivers, limited information on ecological, relevance, (lack of realism)	Comparative physiology, Community ecology
Single driver/community	FOCE, *in situ* pelagic mesocosms (Riebesell et al., 2013; Barry et al., 2014) seeps (Fabricius et al., 2014)	*In situ* removes many laboratory artifacts Community as opposed to species response. Pre-adapted communities (seeps)	Logistically challenging, no information on relative influence of other drivers	Comparative physiology, Community ecology
Single driver/evolution	Acidification/adaptation Schaum and Collins (2014)	Connects plastic and evolutionary responses, specific responses to projected future conditions	No information on relative influence of other drivers; size of experiments limits use to model species (but see Scheinin et al., 2015)	Microevolution
2 or 3 way multiple driver /one species	Warming and acidification (Parker et al., 2009)	Individual versus interactive effects	No information on ecological relevance (lack of realism)	Comparative physiology
4 way multiple driver/one species	Warming, acidification, light and trace metals Xu et al. (2015)	Individual versus interactive effects	Difficult to conduct and also interpret, no information on ecological relevance (lack of realism)	Comparative physiology
Multiple driver/competition experiment	Warming/Acidification Moustaka-Gouni et al. (2016) (2 drivers)	Competition as opposed to single species	Limited information on ecological relevance (lack of realism)	Comparative physiology
Multiple driver/community	Alsterberg et al. (2013)	Direct and indirect effects, synergies and antagonisms	Logistically difficult and resource intensive	Response to IPCC projections Community ecology

Experimental Approach	Examples	Benefits	Disadvantages	Main uses
Multiple driver/evolution	Brennan et al. (2017)	General evolutionary mechanism and limits; connects plastic and evolutionary responses	Logistically challenging and time-intensive, no information on ecological relevance (lack of realism)	Microevolution
Multiple driver/'collapsed design'	Boyd et al. (2015)	Cumulative effects and influence of individual versus interactive effects	No information on ecological relevance (lack of realism)	Reaction norm Response to IPCC projections
Multiple driver/'fractional design'	Gunst and Mason (2009)	Efficient testing of main effects in large multi-driver designs	No intermediate driver levels; frequently lack interaction terms	Identify key drivers in multi-driver factorial designs
Multiple driver/'reduced design'		Cumulative combined effects; Increased power to test hypothesis of interest	No information on ecological relevance (lack of realism)	Reaction norm Response to IPCC projections

TABLE 1. Summary of the main experimental approaches used in multiple driver research, their advantages, disadvantages and which research themes or fora they have mainly been used in. Note many of the research questions posed throughout this review cannot be solved by one single experiment or experimental approach. Scenario-based experiments not only permit more replication (because of fewer treatments and treatment combinations), and hence greater statistical power, within the available resources, and also enable tests of more drivers, in different combinations, and/or at more levels. This is essential for identifying emerging patterns of how drivers interact (e.g. Brennan & Collins, 2015). The benefits of such scenario testing include the development of practical methods to test for multi-driver effects that integrate the modulating effects of interacting drivers, and which can be applied beyond the species-level (i.e. in community-level experimentation). Table from Boyd *et al*. (2018)

1.4 RATIONALE FOR A WEB-BASED BEST PRACTICE GUIDE

The complexity of the natural environment and the number of drivers often prevent the design of a single fully factorial experiment evaluating the impact on a given species or ecosystem. For example, a full factorial experimental design to investigate 4 drivers, with 5 levels per driver, and five replicates would require 3,125 experimental units across 625 treatments (see Table 3 on page 30). To address such complex questions, it is then critical to break them into a suite of simpler experiments. Such simplification requires guidance to ensure that such a strategy follows best practices and have enough statistical power to identify impacts.

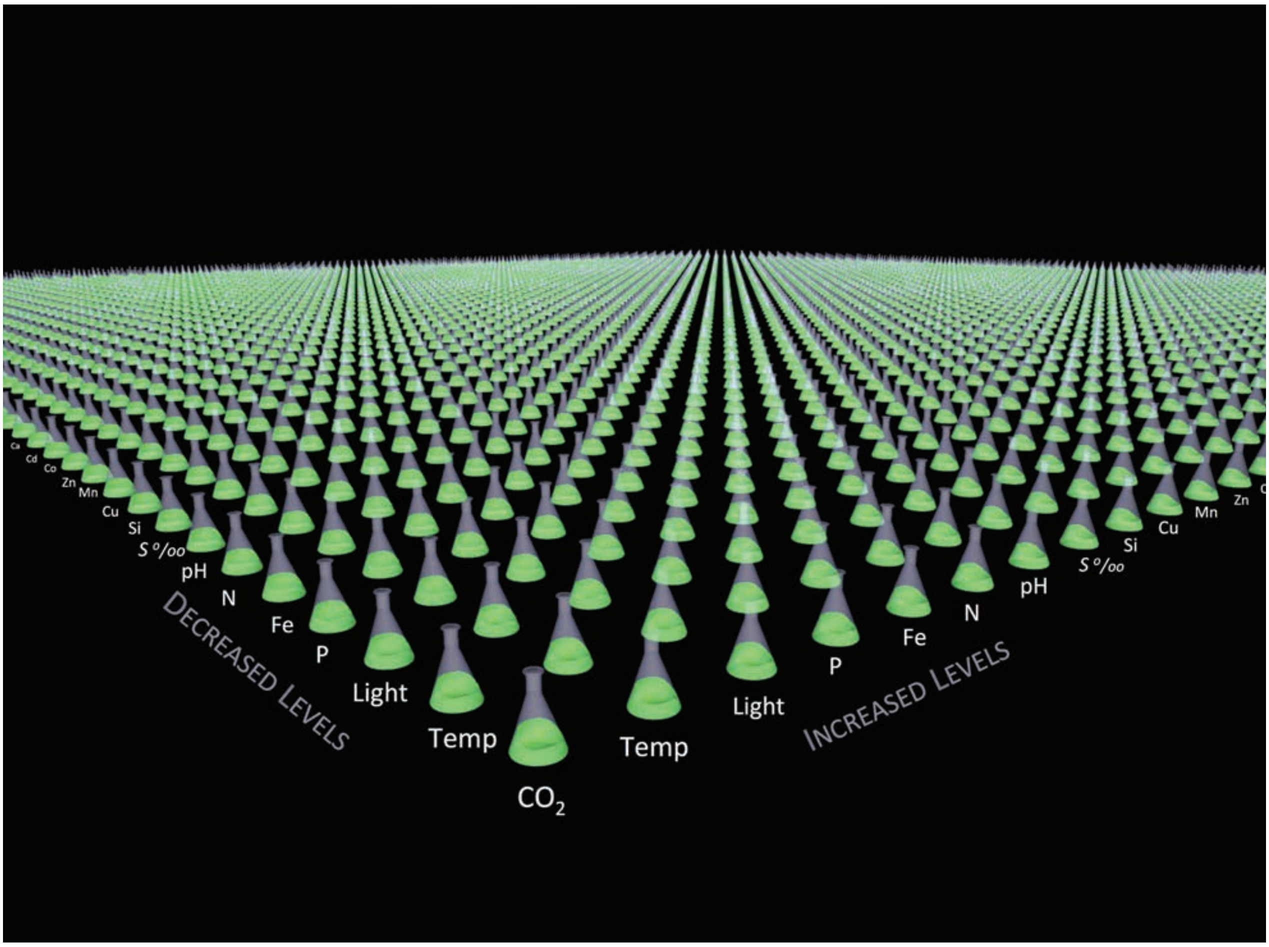


FIGURE 6. A cartoon illustrating the perils of designing an overly ambitious experiment, that requires a large amount of resources, time to run and analyse, and often problems in the interpretation of the data (See the video tutorial by Jon Havenhand within the www-based Best Practice Guide). Cartoon courtesy of Brook Nunn and Keith Holcombe.


EUROPEAN COMMISSION / Research & Innovation / Environment

Guide to best practices for ocean acidification research and data reporting

STUDIES AND REPORTS

EUR 24872 EN

Guide to best practices for ocean acidification research and data reporting

Table of contents

7

FIGURE 7. The cover of the Ocean acidification Best Practice Guide along with a sample section from the table of contents that demonstrates the many aspects that must be considered for such studies.

Best Practice Guides have been written for single drivers such as ocean acidification, resulting in a 258 page publication (Riebesell *et al*., 2011) that detailed the many facets that need to be considered when conducting acidification studies including manipulation experiments – from the fundamentals of carbonate chemistry to which treatment levels to employ, to modelling and data analysis.

In the case of multiple drivers, the following characteristics provided the rationale to use a web-based approach rather than a book:

- Many drivers each with different characteristics
- The driver inventory for each researcher or study subject will differ
- The confounding effects of the interplay between drivers

A web-based best practice guide provides a nimble and flexible approach to this complex research theme. It is also amenable to being readily updated, as is the expectation in this rapidly emerging research discipline.

1.5 NAVIGATING THE BEST PRACTICE GUIDE

The web-based BPG has three distinct components that are linked, with each focussing on a different mode of learning.

The decision support tool is designed to help you get started, and has different entry levels to facilitate giving the right pointers to the newcomer, as well as to those at an intermediate or advanced level.

MEDDLE – Multiple Environmental Driver Design Lab for Experiments – enables you to design and run experiments on a website and hence promotes self-learning and upskilling.

A library of videos enables you to refine your skill-set via a series of topical tutorials by field-leading experts.

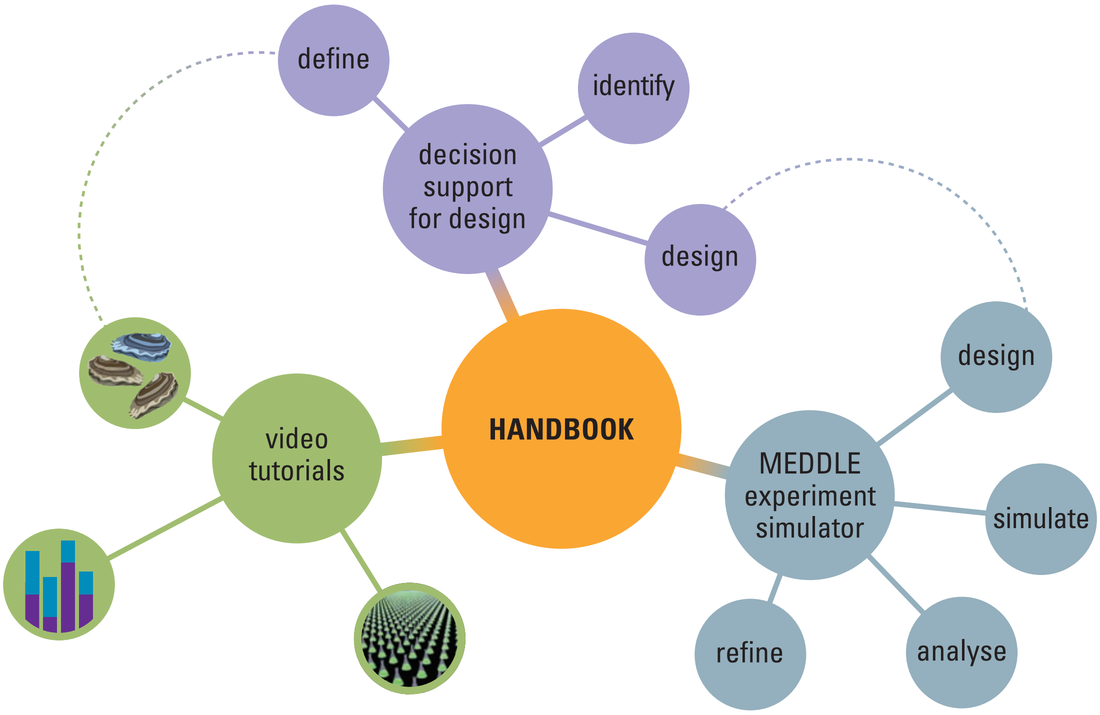


FIGURE 8. This Handbook links the three strands of the Best Practice Guide: (a) a web-based decision support tool, (b) Multiple Environmental Driver Design Lab for Experiments (MEDDLE) simulation software, and (c) a library of video tutorials.

1.5.1 Decision Support Tool

The Decision Support tool takes you through a different stage of the planning process:

(1) defining the research question

(2) identifying responses, drivers and the design

(3) finalising the design.

Together they form a series of three iterative loops designed to systematically step you through a series of decisions needed to arrive an experimental design that you can then use in MEDDLE. The three loops also often multiple entry points depending on the users' level of expertise on this topic.

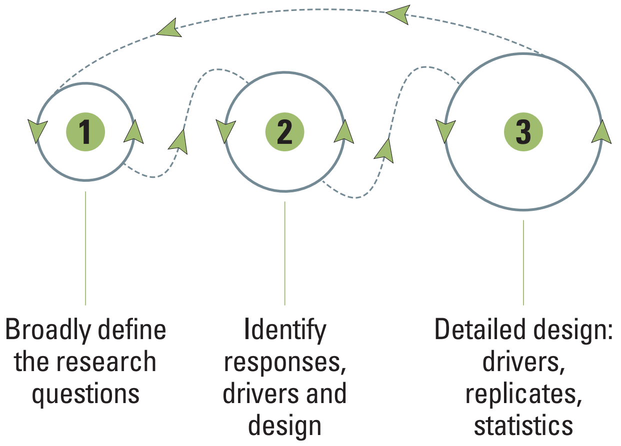
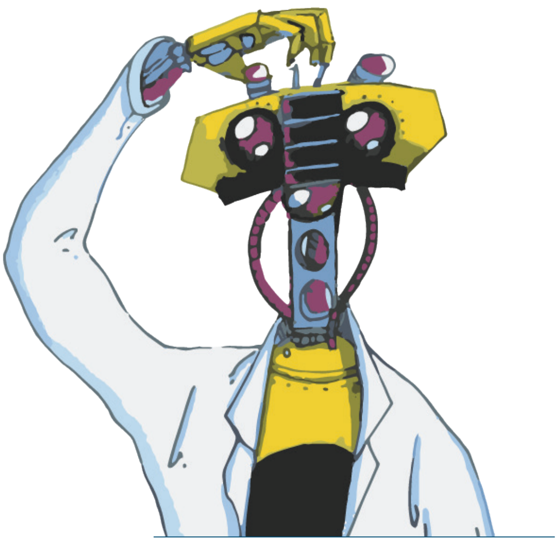


FIGURE 9. The three iterative loops – along with their individual roles – for the decision support tool.

The Decision support tool can be filled in on line or each of the loops can be downloaded as a Word document and filled in. It is envisaged that the completed documents can be discussed with supervisors, within lab group meetings or with mentors. The documents can then be further amended, revised and the initial experimental designs refined in preparation for the next strand of the Best Practice Guide – MEDDLE – Multiple Environmental Driver Design Lab for Experiments.

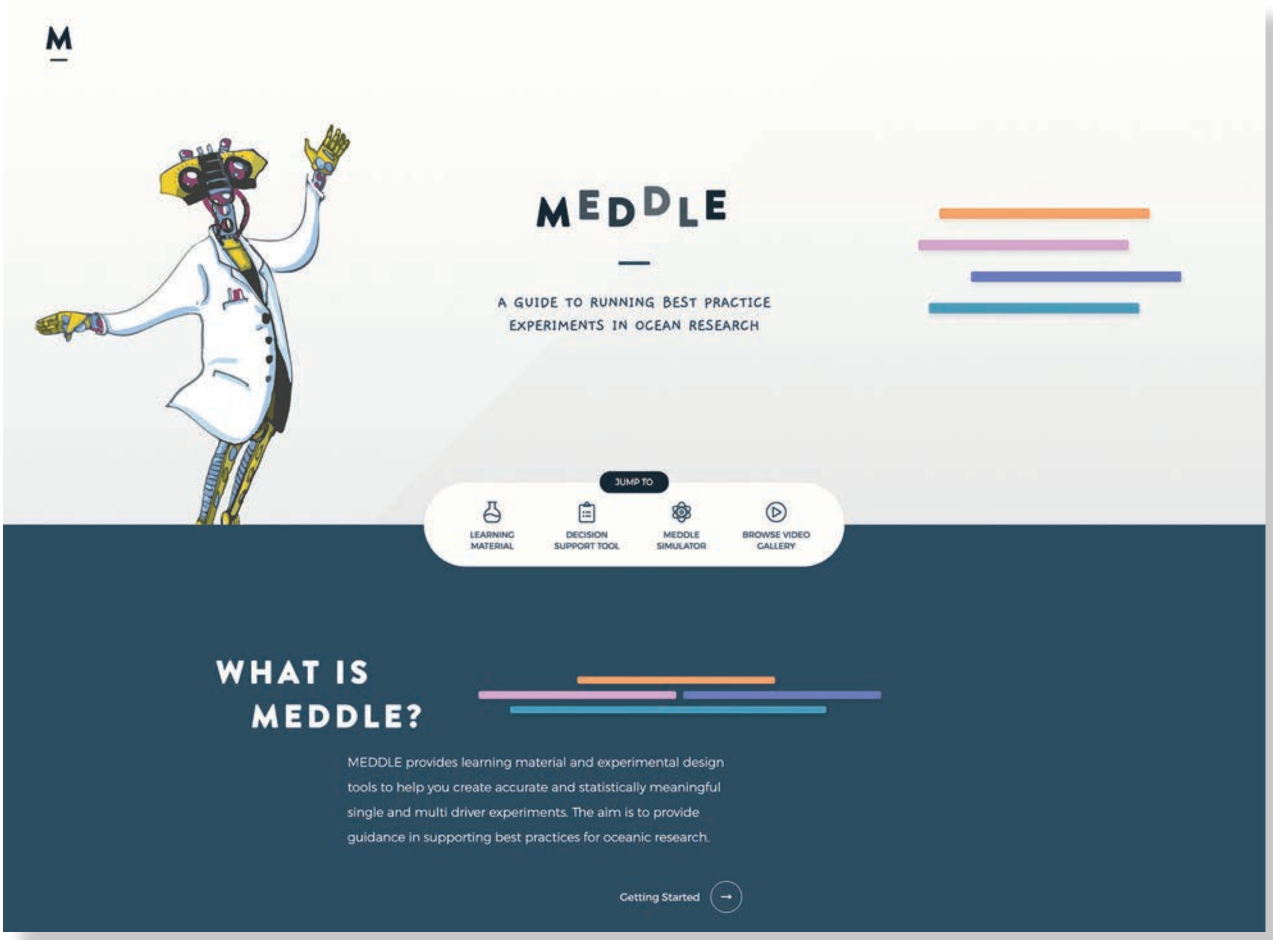


1.5.2 MEDDLE

MEDDLE provides both background learning material and the entry point to the experimental simulator. For the newcomer to multiple driver research this background material can be used to explore the nature of response or affinity curves for different temperatures or for other single drivers such as salinity, or carbon dioxide. Others at more advanced stages may wish to proceed directly to the learning material on multiple drivers. The next step is to become familiar with the simulator where single or multiple driver experiments (such as those featured in Table 1) can be run using a wide range of permutations of treatment levels and replication.

In most cases it is best to commence with a series of single driver experiments to get a handle on the characteristics of each of the drivers within MEDDLE. This mimics the running pilot or preliminary studies which can greatly assist with the step up in logistical, conceptual and analytical skills needed to run multiple driver experiments.

The output file (in csv format) can be readily analysed, and there are some pointers and resources on how to go about data visualisation and statistical analysis. The final step in MEDDLE is Refine – where you can determine if your design is good, adequate (i.e., in need of refinement) or inadequate (in need of major refinement, or re-design).



MEDDLE provides learning material and experimental design tools to help you create accurate and statistically meaningful single and multi driver experiments. The aim is to provide guidance in supporting best practices for oceanic research.

FIGURE 10. A screen shot of the opening page on the www-based guide of MEDDLE.

1.5.3 Video tutorials

The final mode of learning in the www-based BPG is a library of video tutorials which delve into and amplify many of the issues addressed in this Handbook, in the decision support tool, and in the background learning material within MEDDLE.

The tutorials commence with three introductory videos which cover the fundamentals of planning a multiple driver experiment: Driver Inventories, Experimental Design, and Data Analysis. Viewers are then encouraged to advance to more specialised videos including Environmental Realism, Ecology and Evolution, Meta-analyses, and Scenarios vs. mechanisms.

This video series will help you design, carry out, and analyse data from multidriver experiments. Developed by Scientific Committee on Oceanic Research (SCOR) Working Group 149, these

The first three videos provide essential background information for designing and carrying out multiple driver experiments. After watching these introductory videos, move on to the Topical

After watching the "Getting Star with the Topical Tutorials. Let u

Option 1: Full factorial replicated in time

Resource intensive, but good for mechanistic understanding

FIGURE 11. A frame from the video tutorial Experimental Design featuring the biostatistician Professor Jon Havenhand. In this tutorial Jon steps the viewer through a range of experimental design options.

2. CONSTRUCTING A MULTIPLE DRIVER INVENTORY

2.1 INTRODUCTION

Multiple drivers represent a complex matrix of changing ocean properties, and furthermore this matrix may vary from locale to locale. So, before you begin to design your experiment, the first step is to identify which drivers are important in your study region. It is advisable to do this, as drivers which have been widely studied – such as carbon dioxide and pH in ocean acidification – may not be the only ones which are important at your study site(s). Thus, it is important to have an open mind as you construct an inventory of the drivers that can potentially impact marine life in your region. It may be the case that a regional or local driver can exert a significant influence on the biota where you plan to conduct your research. By constructing an inventory, you are taking a holistic approach to the design of multiple drivers. It is likely that you will end up with a combination of global, regional and local drivers.

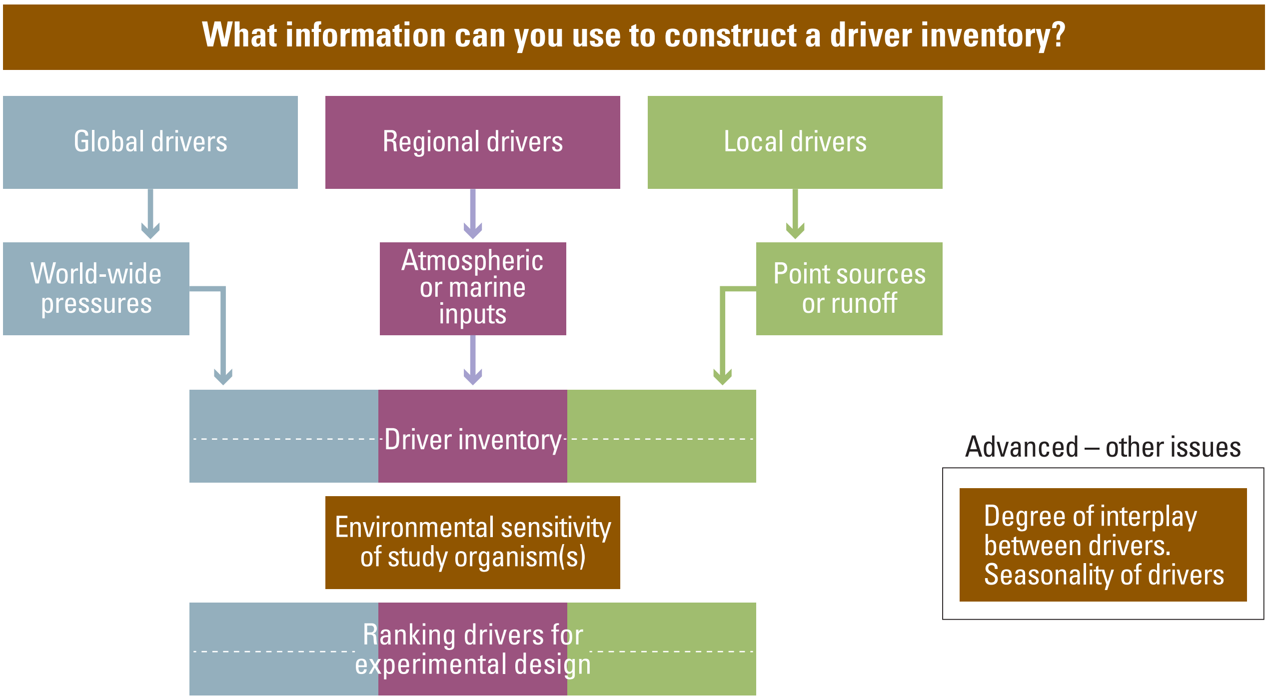


FIGURE 12. Flow chart to assist with constructing a driver inventory for your study subject (organism(s), locale, etc.). This inventory will assist you in ranking the multiple drivers, most relevant for your study subject, as you begin to design your experiment. In the lower right part of the figure are several other issues that more advanced designs may wish to consider.

An example of a very detailed multiple driver inventory is provided in Table 2. The main take home from this inventory is the wide range drivers, some of which are local and some are global, and which can be defined as either acute or chronic threat (see Multiple Drivers within the learning materials for more details on these terms in the context of multiple drivers). It is important to note that this inventory, along with threat and risk assessment was the product of a large team of researchers over a long period of time. Hence, your driver inventory will be less developed than that in Table 2.

Driver	**L**ocal vs **G**lobal cause	**A**cute vs **C**hronic	Risk to Ecosystem
Ocean warming	G	C/A	Very High
Ocean acidification	G	C	Very High
Cyclones/ altered weather patterns	L (G)	A/C	Very High
Illegal fishing and poaching	L	C	Very High
Incidental catch of species of conservation concern	L	A	Very High
Nutrient runoff	L	C/A	Very High
CoTS	L	A	Very High
Sediment runoff	L	C/A	Very High
Ports/Urbanisation (Habitat modification)	L	C	Very High
Sea level rise	G	C	Very High
Pesticide pollution	L (G)	C/A	High
Barriers to flow	L	C	High
Discarded catch	L	A	High
Extraction of predators	L	A/C	High
Dredging (disposal)	L	A/C	High
Marine debris	L (G)	A/C	High
Extraction from spawning aggregations	L	A/C	High
Outbreak of disease	L (G)	A	High

TABLE 2. An advanced form of a multiple driver inventory assembled from a large body of research from Australia's Great Barrier Reef. List of drivers (or pressures/threats) affecting the Great Barrier Reef World Heritage Area, that were rated as 'Very High or High risks to the Great Barrier Reefs Region's ecosystems' by the Great Barrier Reef Marine Park Authority's Outlook Report (from: Uthicke *et al*. 2016). CoTS denotes Crown of Thorns Starfish – a coral predator.

It is highly likely that you will have many more drivers in your inventory than you are able to incorporate into an experiment. For example, for phytoplankton in the polar Southern Ocean the driver inventory might include: irradiance, carbon dioxide, nutrients, trace metals, temperature (i.e., global drivers) and also UV radiation (a regional driver due to the presence of the ozone hole). Hence, a key issue to resolve in conducting an experiment is the need to reconcile the many drivers that may influence the performance of marine life, with those drivers that your study organisms are most sensitive to.

How you rank the drivers in your inventory with respect to the environmental sensitivity of the study organism(s)? (see Griffen *et al*. 2016 for example) A useful place to start is to look at the literature. Have other labs published studies using a single or multiple driver on your study organism(s) and/or in a comparable locale (cold temperature, tropical) to where you will conduct you study or isolate your study organism(s)? It is equally important to search for information for what is known in terms of "mode of action" (see Figure 2a and b). This can be critical for some drivers (e.g. toxicants) and is a cornerstone of the ecotoxicology literature. Can you make use of this prior research to inform your selection of drivers? You may need to supplement this published research with a pilot project running an initial manipulation experiment on a single driver. In particular, this may be the case if there are no publicly available data or findings on your study subject, or on the driver within your inventory.

Once you have constructed the driver, you can move to the next step which is to consider whether there is marked seasonality in any of the drivers. For example, do seasonal patterns of rainfall influence the run-off of nutrients from the land? Information on seasonality helps to identify and explain what aspect(s) of the variability is important (e.g. not the average), but extremes may be more important. Also, how predictable are the seasonal trends (see the following sections). It is also critical for the discussion of what treatment level/ concentrations to use (see the following sections).

A further issue to take into consideration is whether there is evidence from the research literature of interactions between drivers. For example, warming and acidification can interact due to the influence of temperature on gas solubility. Once you have worked through these additional facets of the environmental drivers, it is time to consider how sensitive the subject of your study is to the drivers within your inventory.

In the following sections of Chapter 2 we provide examples of where to access data to ascertain the role of local, regional and global drivers for your area of interest. We next offer insights into how to use mathematical model projections of future environmental change to help select the treatment levels for experiments (when relevant to the design).

2.2 RESOURCES TO INFORM YOUR EXPERIMENTAL DESIGN

You will first want to establish the range of drivers that are relevant to your study subject. Next, it is time to explore and gather the resources you will need to begin to develop your design. These will include:

- Where to find environmental datasets to bring realism to your experimental design
- Accessing model projections to aid the selection of treatment levels in your experiment

At the same time, you might also want to look at what type of biological response will best resolve your research question(s) or hypothesis. Field observational or experimental data along environmental gradients, e.g., those based on 'space-for-time' substitution or transplantation experiments, can be helpful to form hypotheses about niche widths and some of the potentially most and least ecological functions, sensitive species, and physiological responses. In some cases, it may also be useful to look into the geological past (the paleo record) to see if analogues exist for your study subject or the question(s) you are trying to resolve via your experimental design. These resources can be summarised as follows.

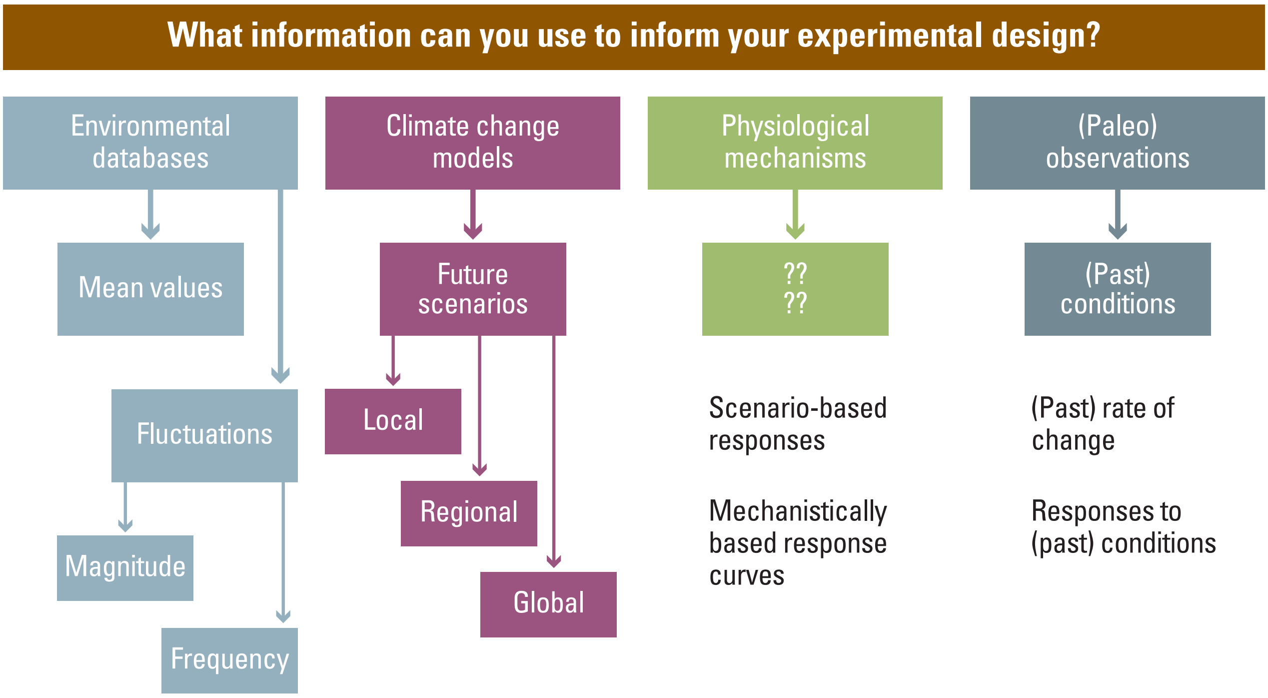


FIGURE 13. A combination of databases, models, laboratory and field studies will help you develop your experimental design. Observations from both the present day and geological past (i.e., paleo observations – which can provide a long-term view for example on ocean acidification – see Hönisch *et al.*, 2012). For insights into the utility of designs using scenario-based versus mechanistic studies see the video tutorial by Sam Dupont on the www site.

2.3 ACCESSING ENVIRONMENTAL DATASETS

There is a wide range of environmental datasets – from local (meteorological station) to regional (national databases) to global (IPCC). Such databases can provide a suite of observations to set up the control treatments in experiments, or to look at patterns in natural fluctuations in marine conditions. It is equally important to first consider and then capture the level of variability (i.e., weather versus climate) that is most pertinent to the scale at which the biology (organisms to communities) will respond. It is essential to look at the environment from the viewpoint of the organism(s) by understanding their niche (the suite of conditions encountered by the organism). It is also invaluable to take behaviour into account – for example a species moving around will encounter different environments, in contrast to a sessile organism (e.g. in the sediment for an infaunal species).

In some cases, if the data you are searching for is not available, you should:

- Carry out some monitoring (to establish an environmental baseline)
- Compare your study site(s) with similar ecosystems/ regions to make the best educated guess.

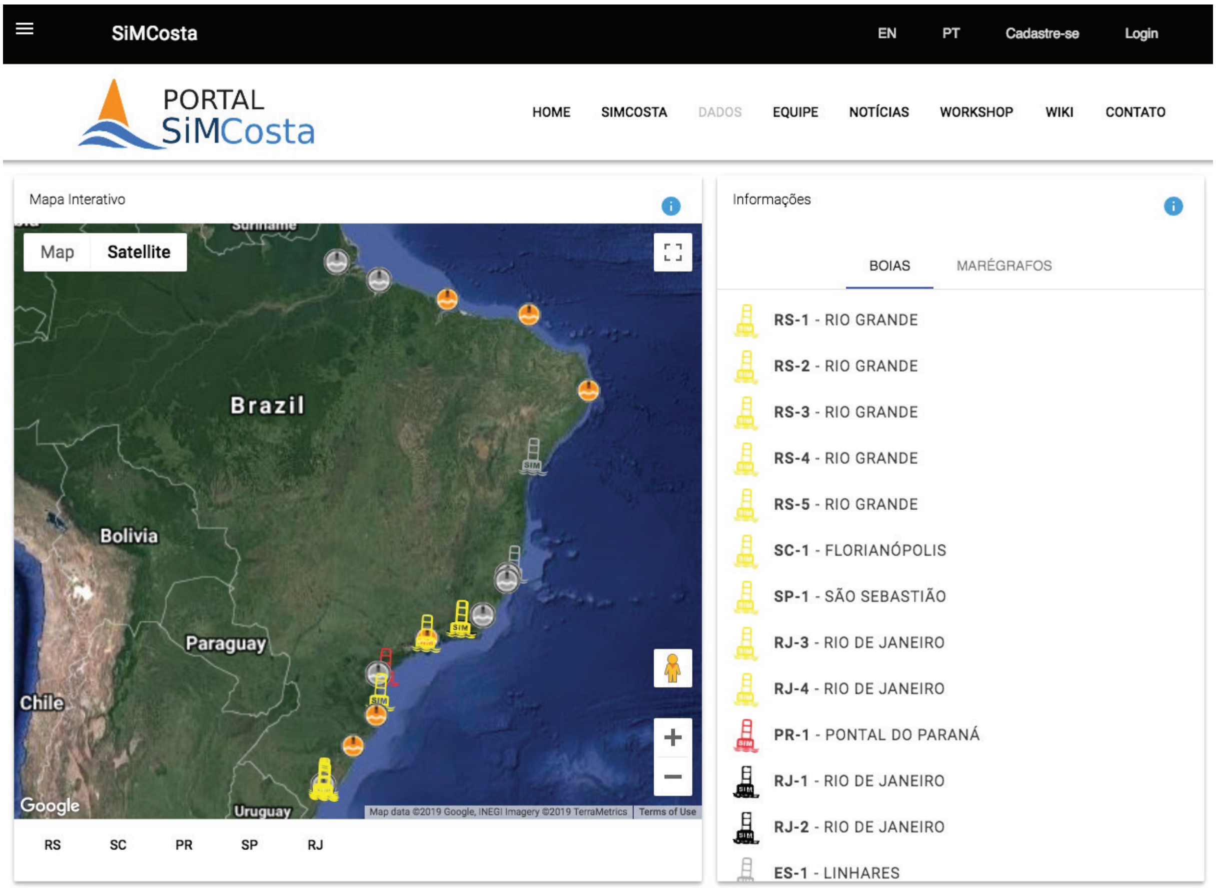
The members of SCOR WG149 come from many different countries, and here we provide two examples of datasets – local and regional – from Brazil and Sweden.

2.3.1 Local

BRAZIL

SiMCosta is a recent initiative of a hand-full of oceanographers in Brazil to establish a net of observatories and tide gauges along the entire coast. We started with 4 meteo-oceanographic buoys around 2014-2016. The site for data distribution is finally up and running (http://www.simcosta.furg.br/). Data are distributed in near real time.

FIGURE 14. Map of Brazil from the SiMCosta website showing the locations of instrumented marine buoys on the eastern seaboard. These clickable links take you to a menu of ocean properties such as temperature and turbidity that can be plotted at a range of temporal scales.



The buoys have a near surface instrumental package that include sensors for temperature, salinity, turbidity and fluorescence by CDOM and chlorophyll, pH and oxygen.

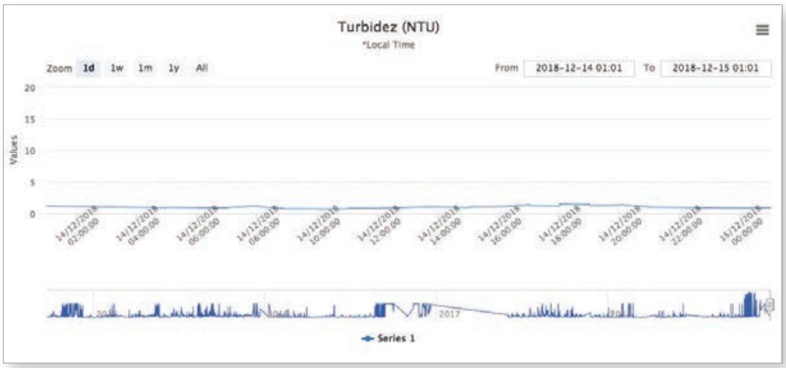
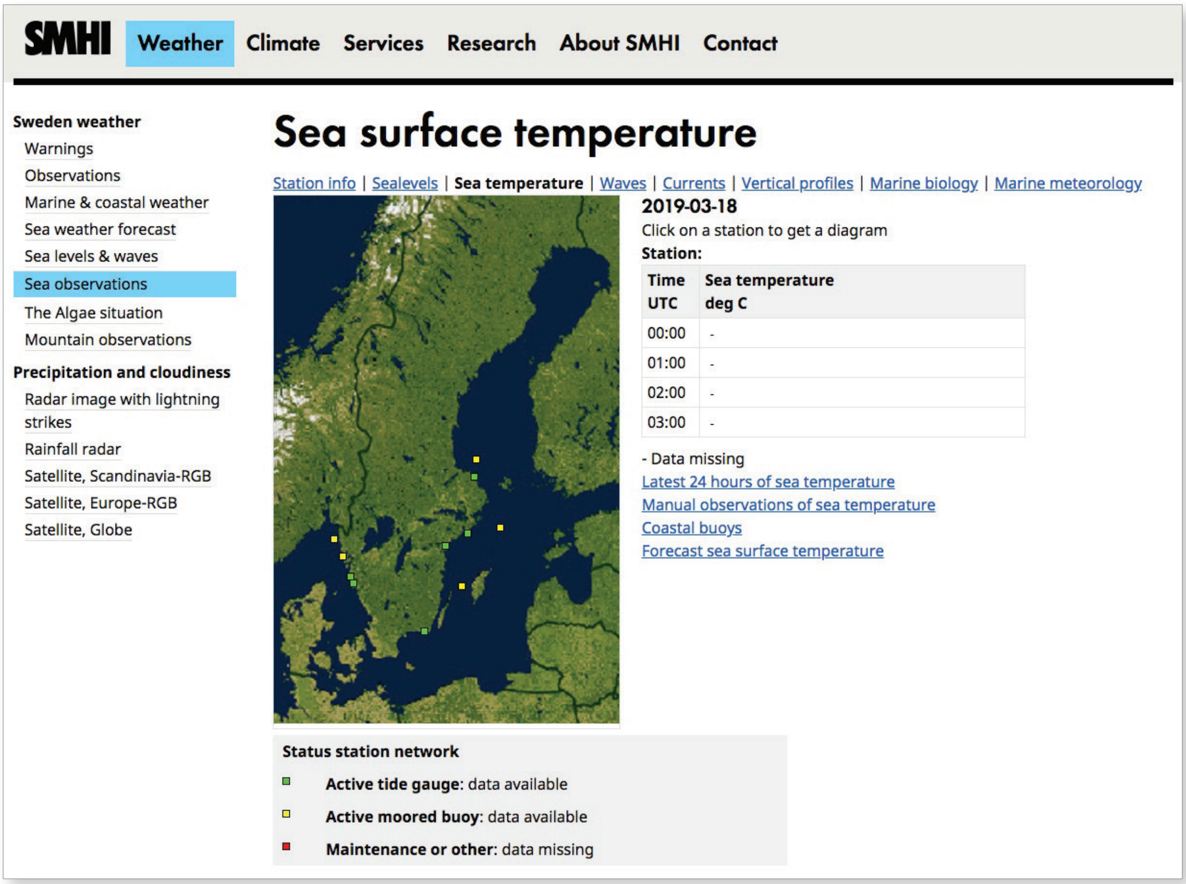


FIGURE 15. An example of a diurnal time-series of water column turbidity from a marine instrumented buoy at São Sebastião (23°50’S, 045°W) from the following website http://www.simcosta.furg.br/

SWEDEN

Swedish Marine Data Archive (Svensk Havs ARKiv – SHARK): https://www.smhi.se/klimatdata/oceanografi/havsmiljodata (in Swedish)

The Swedish Marine Environmental Data Archive in managed by the Swedish Meteorological and Hydrological Institute (SMHI). It contains collated data from hundreds of measurement stations around the Swedish coastline. Temporal resolution of data, and the nature of the data available, varies substantially among stations and time periods, but data on basic variables such as T°, salinity, depth, nutrients, Chlorophyll a, etc. are usually available for all stations. Data for the more offshore stations are collected at regular intervals by SMHI and hence tend to be more complete than data for many inshore stations, which are collected by regional and local agencies on behalf of SMHI. Nonetheless, there are excellent data available. Note the pH data are obtained using standard methods, but with NBS calibrations, and hence may not accurately reflect true SW pH. So, it is important to look at the quality controls for each of these datasets online.



Sea surface temperature

2019-03-18

Click on a station to get a diagram

Station:

Time UTC	Sea temperature deg C
00:00	-
01:00	-
02:00	-
03:00	-

- Data missing

Latest 24 hours of sea temperature

Manual observations of sea temperature

Coastal buoys

Forecast sea surface temperature

Status station network

🟩 **Active tide gauge**: data available

🟨 **Active moored buoy**: data available

🟥 **Maintenance or other**: data missing

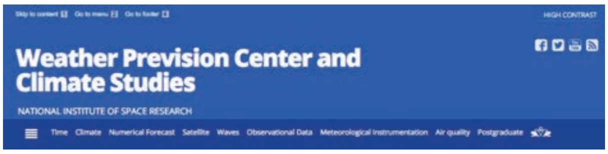
FIGURE 16. An example from the Swedish Marine Data Archive (SHARK) of the location of sea surface temperature time-series observations on buoys. In cases were data are unavailable for your study site you may need to seek other sources (satellite records), proxies (local meteorological stations) or carry out some monitoring to establish an environmental baseline.

2.3.2 Regional

BRAZIL

In Brazil there are many regional sites with useful datasets. Here are some examples for the southern regions of Brazil.

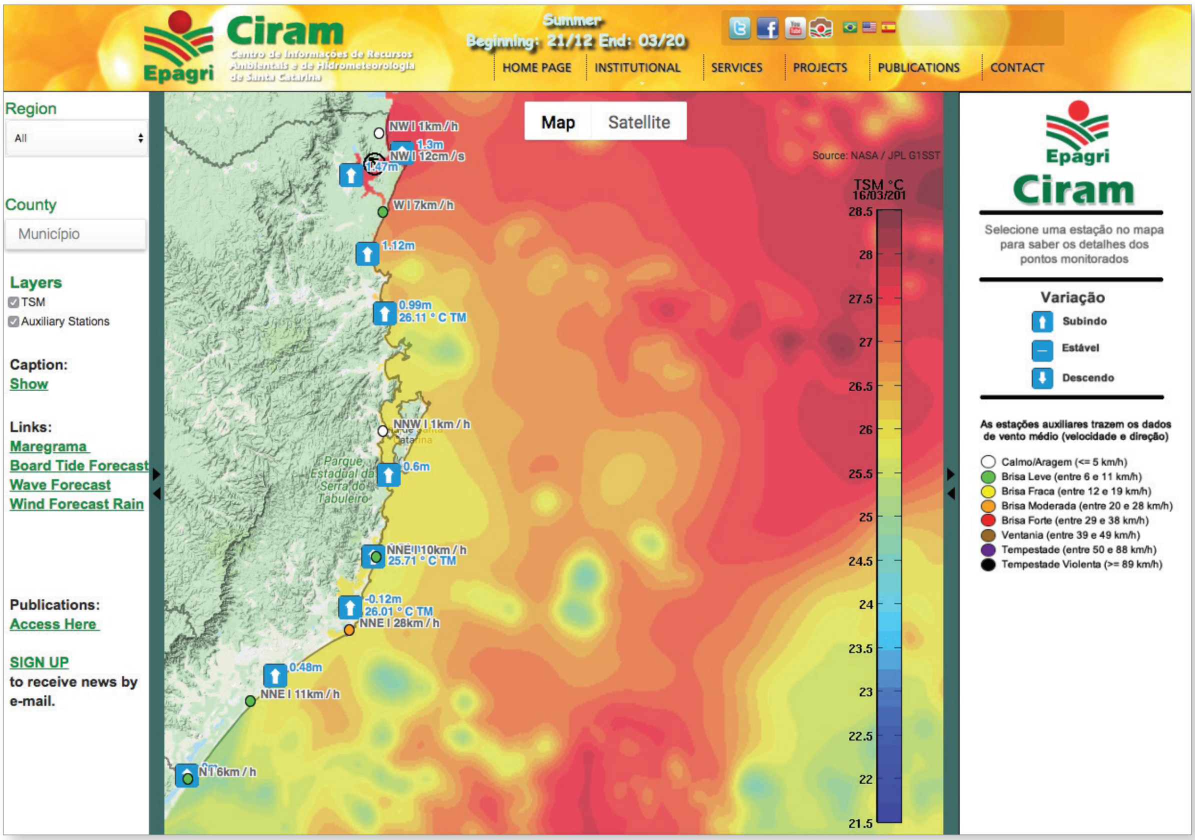
Center of Weather prediction and Climatologic Studies from The National Institute of Space Research (Centro de Previsão de Tempo e Estudos Climáticos – Instituto Nacional De Pesquisas Espaciais): main page in Portuguese but some data links are also in Spanish and English): https://www.cptec.inpe.br/



This is the main site used for weather predictions, funded by the Government, but it also aggregates current and past datasets from observations (balloons, buoys, radars, stations, etc) satellite images and modeling results, including predictions of sea state and precipitation rates. These data will appear as "product options" in many of the other sites listed here. There are some search engines for regional data (by cities), and also educational resources. Projections (Brazilian Global Atmospheric Model; BAM) run over the past 72 hours are kept in the site, which has a graphic interface to produce maps (see example below for accumulated precipitation rates over the past 24 hours). Some apps for cell phones were developed and are available in the site as well.

A regional initiative of the Santa Catarina state (south of Paraná state). Some of the introductory pages are in English and Spanish but the main contents are in Portuguese. This site (http://www.ciram.sc.gov.br/) was created by EPAGRI (Enterprise for research in agriculture for the Santa Catarina state) and it is called Ciran, that stands for "Center for information on environmental resources and hydro-meteorology data", and it was designed to generate and distribute environmental data freely for the general public.

In addition to the core data sets presented in the previous sites, this site has information and prediction on tides along many points of Santa Catarina state coast.


Selecione uma estação no mapa para saber os detalhes dos pontos monitorados

As estações auxiliares trazem os dados de vento médio (velocidade e direção)

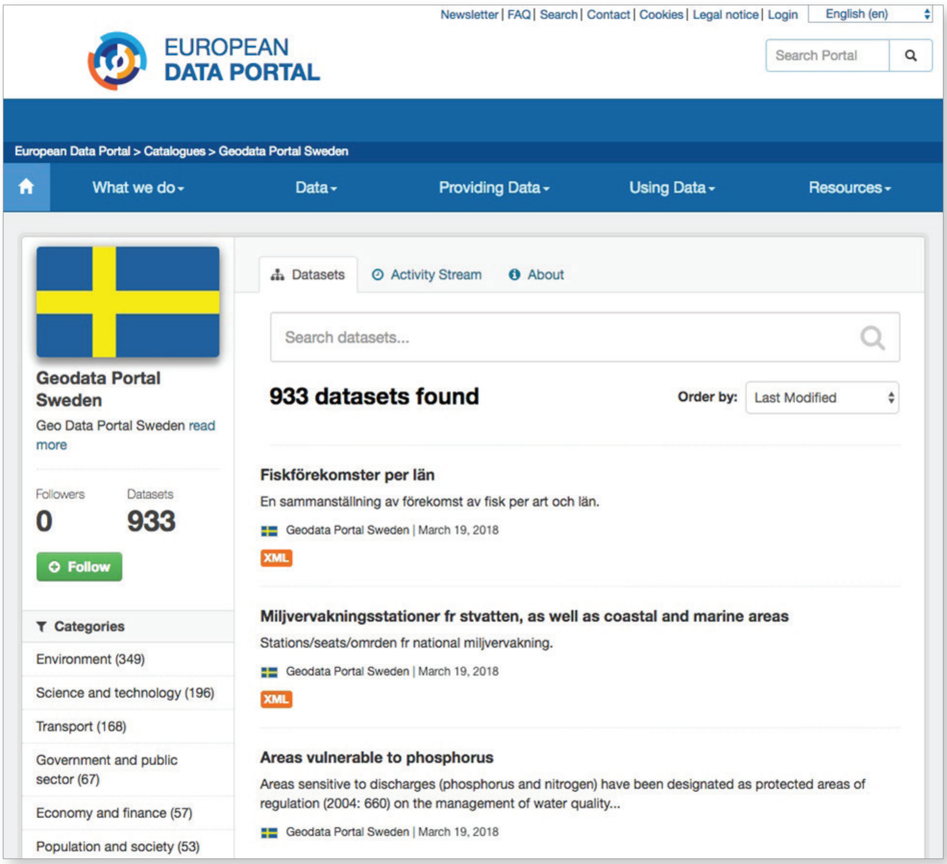
FIGURE 17. A suite of time-series ocean data from three sites that straddle the Santa Catarina state (27.2° S, 50.2° W, south of Paraná state). They provide regional inter-comparisons of ocean properties. This site is located at http://www.ciram.sc.gov.br/

2.3.3 Global

SWEDEN

Swedish Geodata Portal: https://www.europeandataportal.eu/data/en/organization/geodata-portal-sweden

The Swedish Geodata Portal is part of the European Data portal, which provides a meta-data service for all environmental data in Europe. The website contains links to many sources of data, several of which unfortunately result in "404 Page Not Found", however some of the links do work and can be a valuable source of information on (e.g.) distributions and concentrations of environmental pollutants such as heavy metals.



933 datasets found

Fiskförekomster per län

En sammanställning av förekomst av fisk per art och län.

Geodata Portal Sweden | March 19, 2018

Miljvervakningsstationer fr stvatten, as well as coastal and marine areas

Stations/seats/omrden fr national miljvervakning.

Geodata Portal Sweden | March 19, 2018

Areas vulnerable to phosphorus

Areas sensitive to discharges (phosphorus and nitrogen) have been designated as protected areas of regulation (2004: 660) on the management of water quality...

Geodata Portal Sweden | March 19, 2018

FIGURE 18. Screen-shot of the Geodata Portal Sweden that resides within the European data portal. 36 countries (many with coastlines) from Austria to the UK have environmental data on this portal. For Sweden see https://www.europeandataportal.eu/data/en/organization/geodata-portal-sweden

2.4 ACCESSING MODEL PROJECTIONS

Mathematical models and their projections of how ocean properties will be altered under a changing climate provide a wealth of data that can be used to aid in the selection of treatment levels in your experiment. For example, you may intend to design an experiment that considers how scenario-based estimates (for example for the year 2100) will influence your study subject. The following section provides some background information on how models are constructed. It also provides some links to modelling resources and discusses what models can and cannot do.

Using the prior distinction of global, regional and local drivers (see Figure 1a), models are discussed in terms of their resolution, from local through regional to global. The major focus is on global models, followed by regional approaches with few models available on local scales. It is important to bear in mind that that not having access to model projections that relate to your subject of study is not in itself a limitation. In many respects it is just as important to understand what is happening within or without the natural range of environmental variability as there is not a sole control or only one future scenario.

2.4.1 Local modelling

In many cases, the site where your experiment is based may be on a shoreline, in an estuary or an embayment and ideally there would be some model simulations at sufficient resolution to provide projections at a local scale. However, most of these models do not have the resolution required for many locales. There are a few exceptions, such as the model (a 4 km resolution ocean model, validated with >20 observational sites) simulation presented in Figure 19. Even at this high resolution the authors Mongin *et al.* (2016) state that:

> *"The model is too coarse to resolve some of the small-scale water circulation features, such as internal waves, filament and small freshwater plumes. Freshwater footprints are difficult to accurately be represented in a 4-km resolution ocean model. For example, the real freshwater plumes could be thinner than the model grid cell, or could be offset in space and time, which make a comparison with observations at a point in space deceptive."*

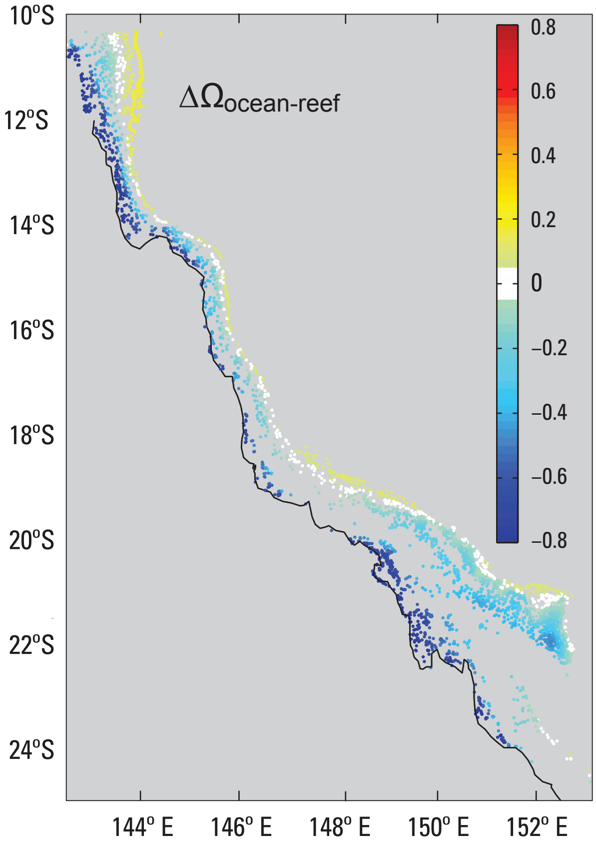


FIGURE 19. Simulated *in situ* aragonite saturation state along the length of the Great Barrier Reef on the Northeastern seaboard of Australia. The map displays the difference between the open ocean value and the value simulated at the reef ($\Delta\Omega$ocean−reef =(Ωocean−Ωreef). Figure from Mongin *et al.* (2016).

In the absence of model simulations at sufficient resolution there are two other approaches that can be used – niche identification or statistical downscaling of lower resolution models.

For niche identification, the present day niche of the study subject needs to be identified. Note, it is also important to take into account biological features. Next, you can use treatments that are within and outside the niche (i.e., stress). Then the interpretation of the data can be done in the light of what is known through models and/or scenarios.

In the case of model downscaling, it requires many resources to carry out and so it is not widely available. It is defined as *"Downscaling is the general name for a procedure that takes information known at large scales to make predictions at small scales."* (Sun *et al*., 2012; Hoar and Nychka, 2008 – see https://gisclimatechange.ucar.edu/sites/default/files/users/Downscaling.pdf for more details).

2.4.2 Global and regional models

These models tend to have coarser resolution – often with 1or 2 degree grid cells. This coarser resolution is due to the computational costs of running complex models that often include an interlinked ocean atmosphere land and cryosphere. Such models may provide useful projections for experiments using study subjects that have cosmopolitan or open ocean distributions.

What are Earth System Models?

Earth System Models (ESMs) are simplified numerical representations of the Earth including the atmosphere, the ocean, the cryosphere and the land. The individual components of the Earth system are connected through fluxes of energy and mass. Processes within and interactions between components are described by mathematical equations. ESMs simulate an internally consistent climate in response to radiative forcing (mainly solar radiation, greenhouse gases and volcanic eruptions) and allow for physical, as well as biogeochemical feedbacks on the latter. https://www.nature.com/scitable/knowledge/library/studying-and-projecting-climate-change-with-earth-103087065

By imposing future evolutions of greenhouse gas concentrations in line with socio-economic development pathways, these models allow to project plausible future conditions for the ocean including ocean circulation and biogeochemistry. https://www.climateurope.eu/a-short-introduction-to-climate-models-cmip-cmip6/

https://www.carbonbrief.org/qa-how-do-climate-models-work

Below there is a discussion on how the output from these models can be used to select treatment levels for some multiple driver experiments. We identify major caveats, point to pitfalls and suggest solutions for avoiding them.

Global versus regional models – what do they each tell us?

In numerical models, processes are represented at the level of individual 3-dimensional boxes. The outcome of processes taking place in an individual box is exchanged with adjacent boxes. All boxes together form the 3 dimensional model grid. The size of boxes or grid cells corresponds to the resolution of the model. The higher its spatial resolution the more detailed its representation of processes. However, high resolution implies a large number of grid cells and thus a high computational cost.

https://www.climate.gov/maps-data/primer/climate-models

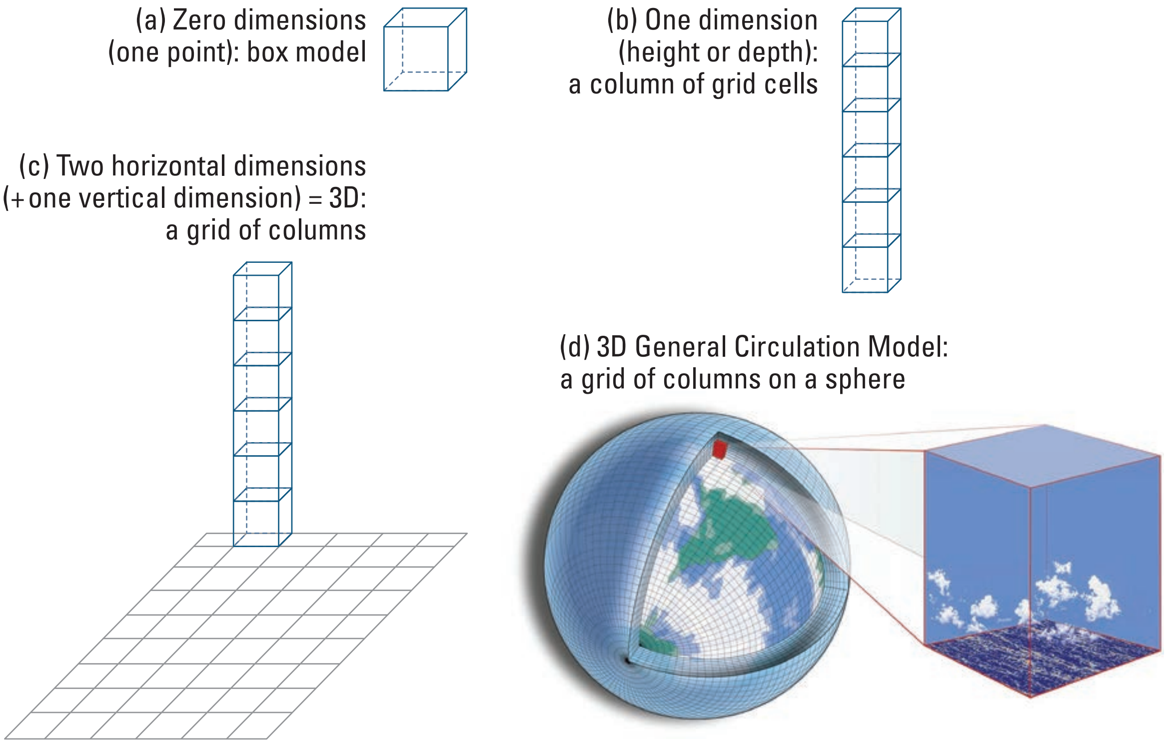


FIGURE 20. Increase in computational costs associated with moving from a single box to a high resolution 3-dimensional grid covering the Earth's atmosphere, ocean, cryosphere and land.

Global ESMs have a high computational burden and climate change projections are still mostly simulated at coarse spatial resolutions of 2° or 1° resolution. This corresponds approximately to 200 or 100 km at the Equator. These resolutions imply that marginal seas, coastal environments and continental shelfs are not adequately resolved in this class of numerical models. **The size of model boxes will be too large to make meaningful inferences of global change impacts on local conditions. In case of a coastal environment, important physical (e.g. tides) and biogeochemical (e.g. early diagenesis) processes will not be represented in the ESM.**

Regional Earth System models (RESMs) can overcome the problem of coarse spatial resolution, but they are still relatively rare. They do not simulate the full Earth, but only a region of it. Most regional climate change assessments are done with circulation-biogeochemical ocean models forced with a climate (the atmospheric conditions) derived from coarse resolution ESMs. Because they do not cover the full globe, regional models require information along lateral boundaries (the limit of the domain covered by the model grid). This information is provided by output from ESMs at regular time intervals (e.g. several hours) to the regional model, which in turn simulates the details of processes in the regional domain. RESMs do not replace ESMs, but provide additional details over a limited spatial domain.

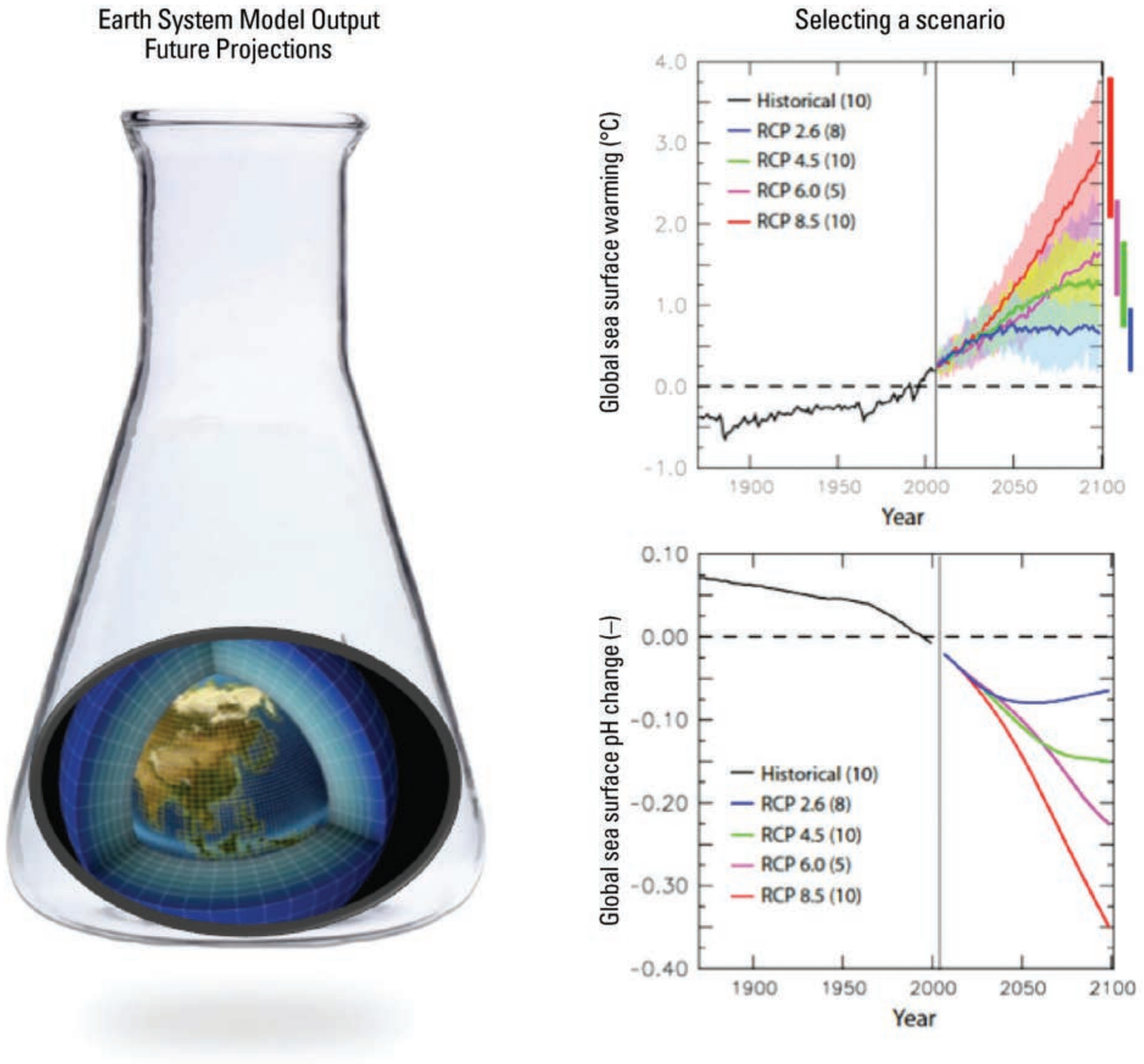


FIGURE 21. Putting the Earth in a test tube: running future climate projections with global Earth system models. Results for different scenarios and from multiple models after Bopp *et al.* (2013).

The preceding paragraphs highlights that no unique and optimal recommendation can be given to experimentalists. The choice of the class of models for identifying treatment levels needs to be done carefully and with reference to the system targeted by the study. As a rule of thumb, one should turn towards global ESMs for experiments targeting open ocean system and regional models for those representing coastal or shelf environments, as well as marginal seas. Finally, output from coupled physical-biogeochemical model systems should be privileged since it will provide dynamically consistent physical and biogeochemical stressors distributions.

Where do we get scenarios from?

Model output for scenarios can be assessed through any Earth System Grid Federation (ESGF) node. Below, in Figure 21 is an illustrative example from the German node. An account is required and it can be created via the link on the page web. Under 'search data', the user in invited to select among different project: CMIP5 (Coupled Model Intercomparison Project Phase 5) or CORDEX (Coordinated Regional Climate Downscaling Experiment). CORDEX provides atmospheric forcing only albeit at high spatial resolution.

ESGF Node at DKRZ

Welcome to the DKRZ ESGF-CoG Node

As a national HPC center, the German Climate Computing Centre (DKRZ) is part of the European IS-ENES Infrastructure. It provides high performance computers, high capacity data storage and data management for German climate research. In addition DKRZ's specialists offer a range of services and support covering modeling and programming as well as data dissemination and long-term archival.

The Earth System Grid Federation (ESGF) maintains a global system of federated data centers that allow access to the largest archive of climate data world-wide.

Start typing, or use the 'Delete' key to show all available tags.

ESGF-DKRZ Tags: *None*

Search Data

The following projects require a ESGF Account (create account) and some also require a Group Registration (see links below) to access their data.

- Search all Projects
- **CMIP6:** Coupled Model Intercomparison Project Phase 6
  - CMIP6 Data Search
- **CMIP5:** Coupled Model Intercomparison Project Phase 5
  - Group Registration: CMIP5 Research CMIP5 Commercial
  - CMIP5 Data Search
- **CORDEX:** Coordinated Regional Climate Downscaling Experiment
  - Group Registration: CORDEX Research CORDEX Commercial
  - CORDEX Data Search
- **Obs4MIPs:** Observations for Climate Model Intercomparisons
  - Group Registration: CMIP5 Research
  - Obs4MIPs Data Search
- **MPI-GE:** MPI Grand Ensemble
  - Group Registration: CC BY-SA 4.0
  - MPI-GE Data Search
- **ISI-MIP:** Inter-Sectoral Impact Model Intercomparison Project
  - Group Registration: ISI-MIP Research ISI-MIP Unrestricted

User help and related links

For technical support please have a look at the User Tutorials and ESGF FAQ.

If you can not find what you are looking for please contact the ESGF user support: esgf-user@lists.llnl.gov

Expert users may also want to use the ESGF Search RESTful API: http://esgf-data.dkrz.de/esg-search/search?

FIGURE 22. An example from the German Climate Computing Centre (DKRZ) of high capacity data storage and data management for climate research. DKRZ also can provide services and support covering modeling and programming as well as data dissemination and long-term archival. https://esgf-data.dkrz.de/projects/esgf-dkrz/

The CMIP5 project should be selected if searching for ocean variables. Model output is grouped in different categories called *Realm* with *ocean* (ocean physics), *ocnBgchem* (biogeochemistry) and *seaice* hosting tracers of interest to the marine scientist. *Variable Long Name* informs on the variables hidden behind the naming convention (*Variable*). Scenarios (rcp26, rcp45, rcp60, rcp85) are found under *Experiment*.

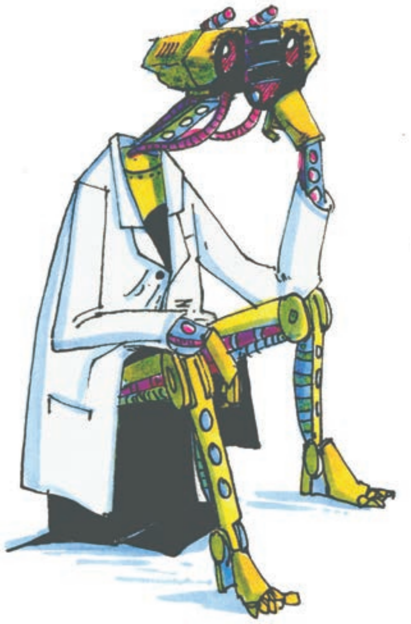
https://www.climateurope.eu/a-short-introduction-to-climate-models-cmip-cmip6/

What do global and regional models project well?

Models are only simplified representations of the real world and model output will always differ from observations. A variety of causes contribute to model bias: differences in spatial-temporal scales models and observations, poorly parameterized or missing processes in models etc. As mentioned above, ESMs have their own climate dynamics. These models reproduce modes of climate variability (e.g. ENSO) in a statistical sense, but not with a precise calendar correspondence. The skill of a multi-model mean is often better than that of any individual model (error compensation) and it should be preferred.

What models do not yet project?

Processes and interactions between ocean physics and marine life occurring at meso- (i.e., <100 km horizontal scales and < 1 month) or submesoscale (very fine spatial scales of <10 km and < 1 week) are not represented in global models and only limited to high resolution regional model systems. Models do in general not yet include daily cycles of biogeochemical variables (e.g. pH, temperature). The interaction between individual drivers and biota is only represented in a very crude way and no acclimation or adaptation is allowed.



3. PARAMETER MANIPULATION

Now that you have an inventory of the driver(s) you wish to investigate, and have consulted relevant databases, literature, and model projections to identify what levels of each driver you might use, it is time to start thinking about how to design your experiment(s). Specifically, there is a need to identify how many experimental units are required (these would be the culture flasks, tanks, or mesocosms to which you would apply to treatments[1]), and how these will be distributed among the different drivers, driver levels, and replicates (i.e., repeat experimental units within each treatment). As will become evident, this can be problematic – and many experimentalists go through the process of testing, and changing, the design of their experiment (sometimes even changing the design of their experimental units) to accommodate the desired design.

Photo below: Monitoring carbon dioxide settings during a multi-stressor (CO_2 and irradiance) phytoplankton experiment in a thermostated water bath. In this experimental set-up eight 85 ml test-tubes are immersed, each independently illuminated by an array of cool white LEDs set at specific intensity and timing and each bubbled by humidified gas of any composition. Image courtesy of the Passow lab (UCSB).

---

[1] Note: a treatment is a combination of the levels of the driver(s) applied to the experimental units. Responses of replicate organisms within an experimental unit are not-independent of the others in the same unit and not the correct analysis unit when analysing the effects of the treatments. Responses of replicate individuals within each experimental unit can be used as technical replicates to compare variability among experimental units within any one treatment, but only at this level: for analysing responses to the treatments themselves we need to use the mean value (or other summary metric) of responses of replicate organisms within each experimental unit.

3.1 CONTROLS AND TREATMENT LEVELS

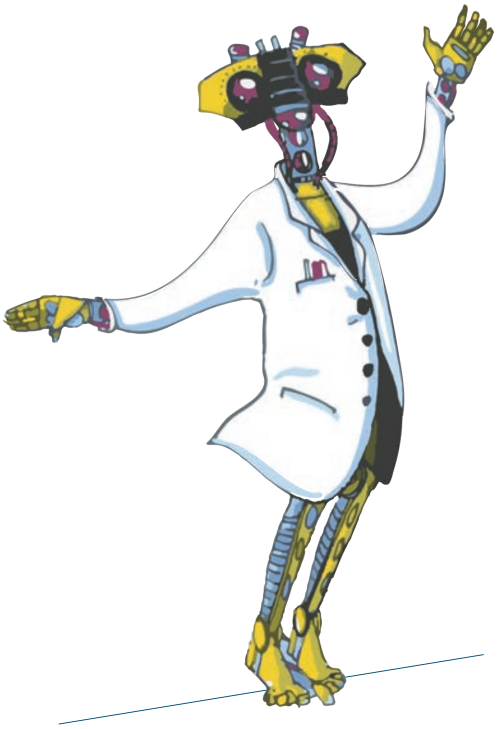
An experimental control is a combination of drivers at given levels (hereafter "control treatment") designed to represent a baseline for comparison with other treatments. Typically, responses to the manipulation of any driver(s) would be compared to those of the "control", allowing experiments to minimize the effects of variables other than the driver(s) of interest. In studies of global ocean change, control treatments are often combinations of drivers that simulate the present-day, or even pre-industrial, environment.

Setting driver levels is a key part of designing your experiment and can influence both the value and relevance of your results. For example, recommendations presented in the Guide to Best Practices for Research on Ocean Acidification (Riebesell *et al*., 2010) greatly facilitated the comparison of responses to standard levels of ocean acidification (Kroeker *et al*., 2013). Such efforts to harmonise experimental treatments are controversial because they are not always relevant to the question of interest. Nonetheless, it's worth considering whether or not this is something that should be included in the design of your experiment in the interests of cross-comparison with other studies (see Section 4 for other examples).

Key issues in choosing driver levels for the drivers of interest include:

- are you interested in environmental relevance of your experiment?
  - if so, do your treatment-levels encompass the range of current, and future variation?
  - will you use the daily/monthly/annual mean, median, or variance as your treatment-level?
  - will your experiment include natural variance as a treatment itself (e.g. simulating diurnal pH fluctuations, or seasonal heatwaves)?
- do you have enough treatment-levels to detect if responses are non-linear? (is this important for this driver?)
- are there agreed "standard" treatment-levels (e.g. positive and negative controls) that you should include?

Setting levels for a range of drivers should take into account the annual range, mean and median, and projected change. Table 3 provides information that may help with the logistics of the design. Setting the treatment levels must be done with care, as illustrated by Figure 23.



Number of factors (i.e. drivers)	Replicates per treatment	Levels per factor	Total number of experimental units			
			Full factorial	Major vectors	Scenario	Collapsed factorial
3	2	2	16	10	4	8
		4	128	26	8	
		6	432	42	12	
		8	1024	58	16	
		12	3456	90	24	
3	3	2	24	15	6	12
		4	192	39	12	
		6	648	63	18	
		8	1536	87	24	
		10	3000	111	30	
		12	5184	135	36	
3	4	2	32	20	8	16
		4	256	52	16	
		6	864	84	24	
		8	2048	116	32	
		12	6912	180	48	
3	6	2	48	30	12	24
		4	384	78	24	
		6	1296	126	36	
		8	3072	174	48	
		12	10368	270	72	

TABLE 3. Calculation of the total number of experimental units based on number of drivers, replicates and treatment levels. For a three driver study, alternative designs (such as scenarios, major vectors and collapsed factorials) described in the video tutorials may be tractable where the full factorial is not. Importantly, they may answer questions of interest just as well as a full factorial experiment would.

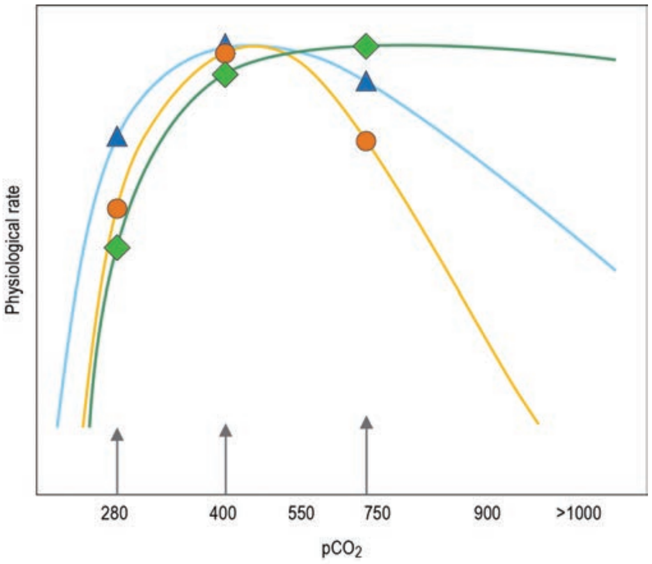


FIGURE 23. Three hypothetical affinity curves for CO$_2$ and a physiological rate process overlaid with different controls and treatment levels to illustrate the importance of selecting each of these carefully. (Figure from Boyd *et al.*, 2018).

3.2 REPLICATION vs. DRIVER LEVELS vs. DRIVERS

Multi-driver experiments can be challenging to run because manipulating the different levels of each driver in the multiple sets of treatments can be technically difficult, often requiring substantial infrastructure. Even if you can overcome these technical challenges, combining several levels of multiple drivers can rapidly generate large numbers of treatments that outstrip the available resources[1] (see Figure 24, Table 3, and (Boyd *et al.*, 2018)).

For a simple one-driver experiment it may be appropriate to use a regression design that has many levels of the driver of interest and no (or very little) replication. Such designs not only provide the possibility of being able to identify a mathematical function that describes the response, which can be used subsequently in models, but can also be statistically very powerful – especially if you only have a limited number of experimental units. (see e.g. (Cottingham *et al.*, 2005)). Such regression designs can even be valuable in two-driver experiments in which there are only two or three levels of the second driver.

Even for multiple drivers an experimentalist may want to use several levels of each driver so that they can see whether responses are linear, and (if not) what form they take. Identifying the different options in trading-off the logistic problems this can create against the information that an experimentalist would like to obtain is central to this Guide – and the reason that the SCOR WG149 created the MEDDLE experimental simulation environment.

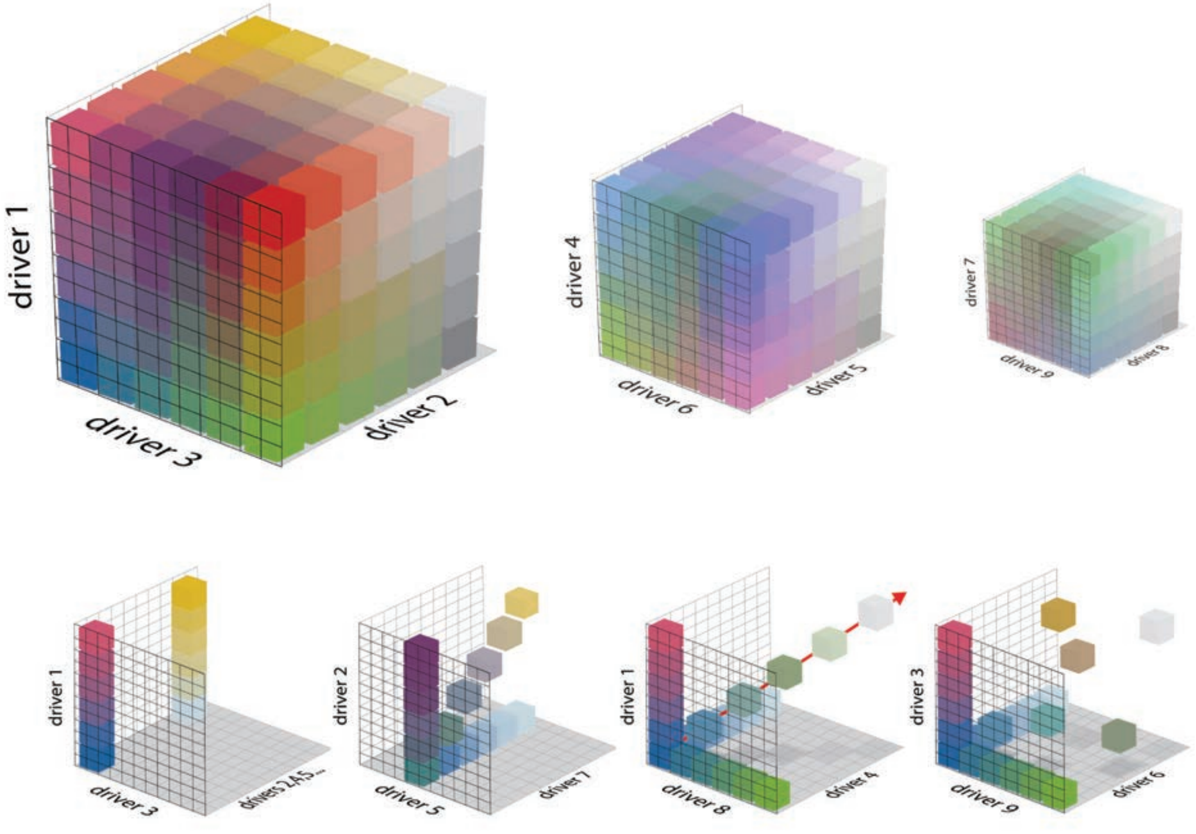


FIGURE 24. Identification of an idealized full-factorial design defining all of the drivers (experimental treatments, here illustrated for three factors) and the range of interest for each one. Next identify the most relevant subset and levels of drivers, and combinations thereof, to create a reduced or collapsed design that best addresses the question(s) of interest (Boyd *et al.*, 2015; Gunst & Mason, 2009). Figure from Boyd *et al.* (2018).

---

[1] Note: a treatment is a combination of the levels of the driver(s) applied to the experimental units. Responses of replicate organisms within an experimental unit are not-independent of the others in the same unit and not the correct analysis unit when analysing the effects of the treatments. Responses of replicate individuals within each experimental unit can be used as technical replicates to compare variability among experimental units within any one treatment, but only at this level: for analysing responses to the treatments themselves we need to use the mean value (or other summary metric) of responses of replicate organisms within each experimental unit.

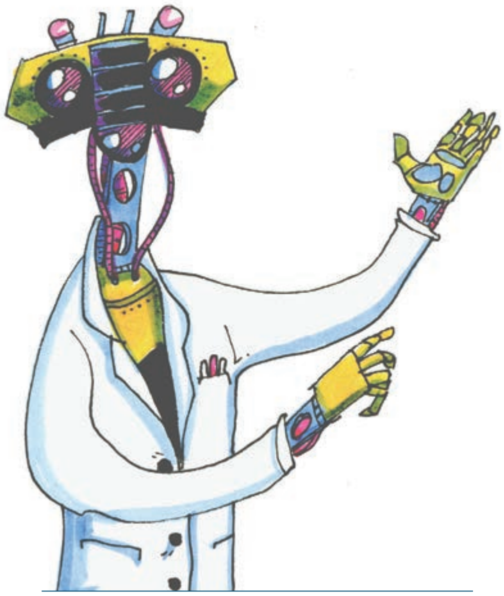
3.3 EXPERIMENTAL DESIGN, ANALYSIS, AND REFINING RESULTS FROM MEDDLE

The MEDDLE Simulator lets the user run trial experiments in silico so that they can test different experimental designs, get sample response data from each of those designs, and then analyse those data and compare results from the different designs to see which might best meet the user's needs. Hopefully this will save users time and money before you get to the laboratory or field.

Briefly, MEDDLE lets the user select up to three drivers, for each of which they can select multiple levels (more than the user is ever to likely be able to use in an experiment), and up to five replicates. You can specify how "noisy" your data are – i.e. how much variance there will be among replicates – and MEDDLE will also now and again randomly drop replicates, simulating the real-world experience of losing a sample or a reading. Depending on the software package used, the "export.csv" file may need reformatting[2].

Identifying how to analyse the results of the experiment – i.e. what graphical and/or statistical tools to use – is an essential step in designing your experiment. (The user will not want to invest a lot of time and effort only to find out they cannot analyse their experiment in the way they had thought!). Fortunately, MEDDLE lets you simulate this process before spending valuable resources in the lab.

Some users will be relatively new to this process, while others will already know how best to analyse their results. So some users will be able to readily carry out their analysis. For any users who are not sure, below are some suggestions to help structure the data analysis. These are neither exhaustive nor prescriptive, rather they are intended to stimulate further thought and inquiry. In all cases, if you haven't already done so, it is suggested that users familiarise themselves with at least one of the many excellent texts on experimental design and analysis (e.g. (Quinn and Keough, 2002, Logan, 2010, Underwood, 1997, Zar, 2013)).



---

[2] The format of the "export.csv" file is not readily readable by most software packages and you'll need to reformat the file.

First Steps visual inspection of the data is an essential first step. MEDDLE provides some basic plots (e.g. Figure 25a), however the user can use the “Export” function to generate an “export.csv” file that can then be read into the user’s favourite software package to generate other plots such as BoxPlots (Figure 25b), simple linear regressions (not shown here because the data in Fig. 25a are curvilinear), or LOESS curves and 95% CIs (Figure 25c).

Intermediate Once the user has inspected the data, they may want to run some basic statistical analysis. In MEDDLE all the drivers are always “fixed factors”[3], the designs are fully factorial, and the data are drawn from a normal distribution (although the user should check that the sample they have obtained from MEDDLE is normally distributed!). In this circumstance you might want to run a simple ANOVA in your favourite software package (see, e.g.: http://rtutorialseries.blogspot.com/2011/01/r-tutorial-series-one-way-anova-with.html).

Advanced For more complex designs, e.g. those including random factors, and/or non-factorial combinations of multiple drivers, the user will need more complex analysis techniques. This is not the place to rehearse the pros- and cons- of those different techniques, and it is recommended that the user consults their favourite experimental analysis text book or ‘pet’ statistician. Examples of the sorts of analysis the user might think about applying to MEDDLE output include multivariate regression (surface fitting) or multi-factor linear models (ANOVA or Mixed-Effects models), e.g.:

https://www.statmethods.net/stats/regression.html

http://rcompanion.org/handbook/G_09.html

http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf

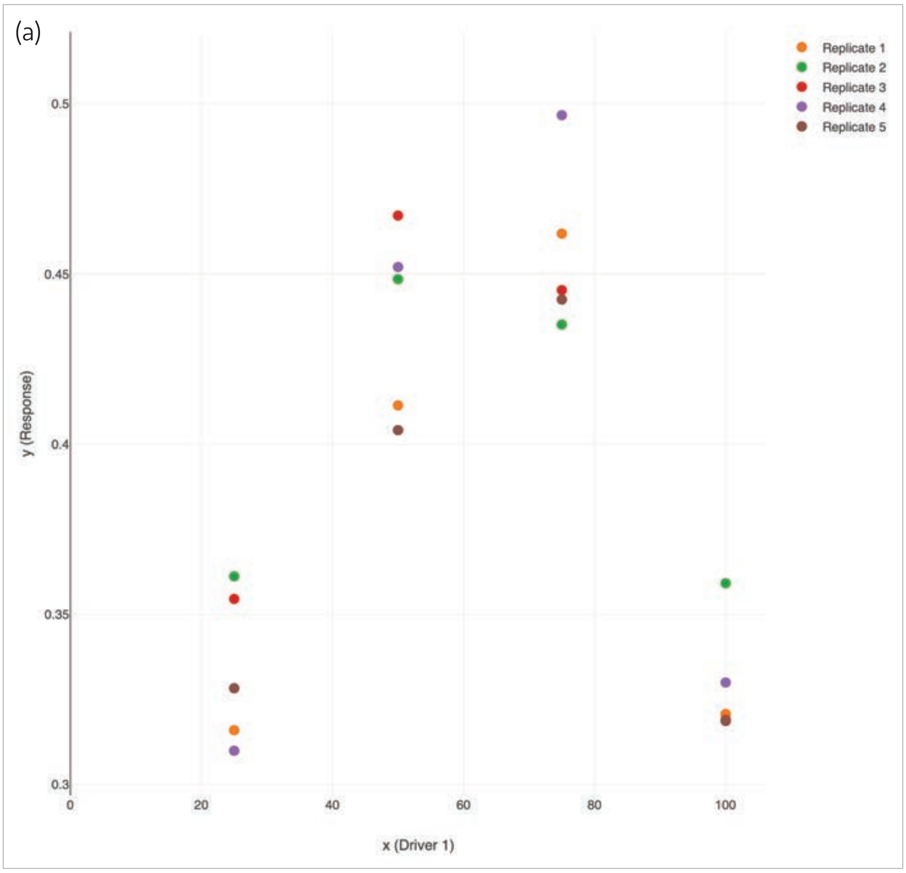


FIGURE 25(a).
An example of a 2-D plot output from MEDDLE.

---

[3] In statistical models, factors can be fixed or random. If you’ve selected the levels of a factor (e.g. different temperatures) then this factor is typically “fixed” (you chose those levels for a reason and would probably choose the same, or similar, levels in a repeat experiment). Conversely, if you randomly sample the levels of a factor from a population (e.g. different populations or locations) then the factor is “random” (you’d chose different samples in any future repeat study).

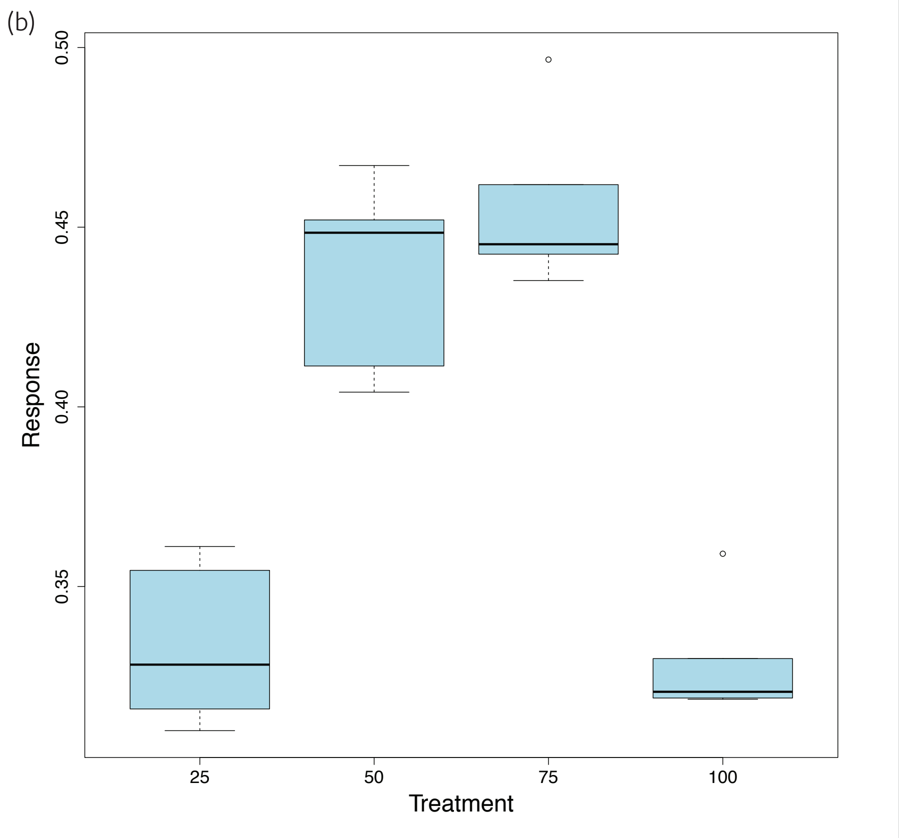


FIGURE 25(b).
A BoxPlot of data in panel A (plot produced in R; see, e.g: https://www.statmethods.net/graphs/boxplot.html)

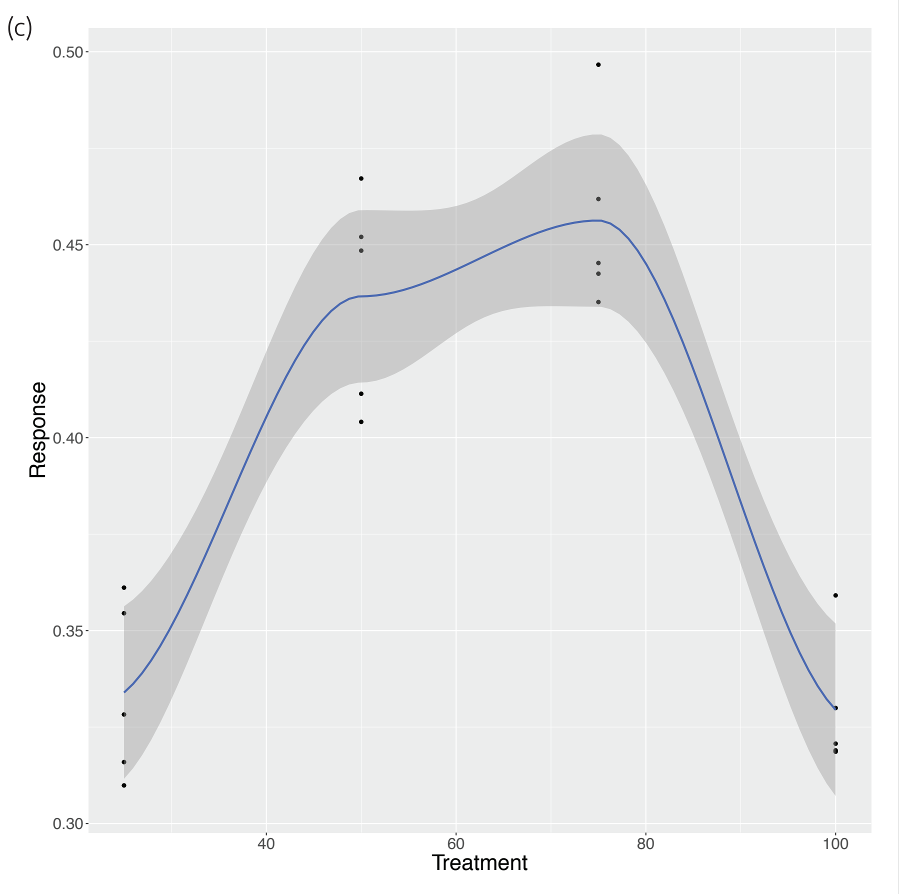
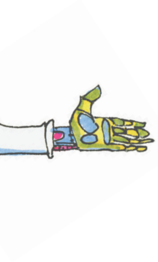


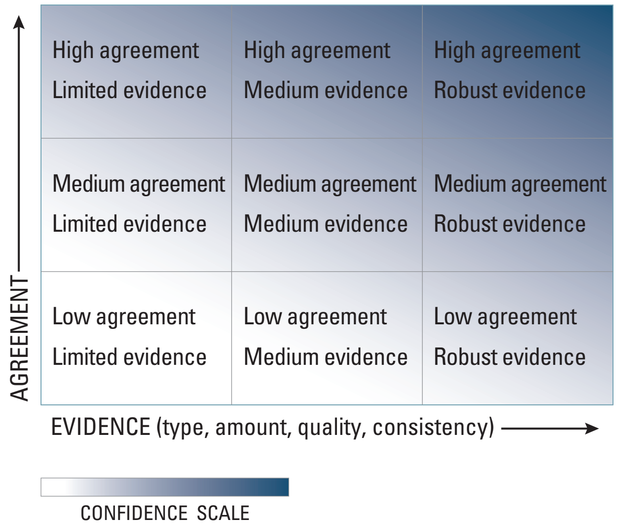
FIGURE 25(c).
A Locally Weighted Smoothing (LOESS) curve with 95% CI for data in panel A (plot produced in R; see, e.g.: http://www.sthda.com/english/wiki/print.php?id=188)
LOESS curves can work well when replication is low but there are many treatment levels.



4. DATA SYNTHESIS AND MOVING BEYOND THE BEST PRACTICE GUIDE

Research into ocean acidification over a decade or more has demonstrated the benefits of a Best Practice Guide to advance a research communities' collective understanding of such a complex topic. It is hoped that the www-based multiple driver Best Practice Guide will also help to advance research by providing insights into the principles of experimental design to resolve questions around the subject under study.

At some point the value of the datasets you have obtained and published from your single and multiple driver experiments can be further extended by making them available for data syntheses such as meta-analysis (Kroeker *et al.*, 2010, and see the video tutorial on this topic) and/or having them become part of wider international synthetic activities such as the IPCC Assessment Report (AR) cycles. The IPCC reports play a key role in attempting to bring together published trends over 5-6 year periods and to seek a consensus viewpoint using a combination of evidence and agreement statements (Figure 26 from Mastrandrea *et al.* 2010). The ability to have studies that are readily comparable with others helps with this process.



High agreement Limited evidence	High agreement Medium evidence	High agreement Robust evidence
Medium agreement Limited evidence	Medium agreement Medium evidence	Medium agreement Robust evidence
Low agreement Limited evidence	Low agreement Medium evidence	Low agreement Robust evidence

FIGURE 26. A depiction of evidence and agreement statements and their relationship to confidence. Confidence increases towards the top-right corner as suggested by the increasing strength of shading. Generally, evidence is most robust when there are multiple, consistent independent lines of high-quality evidence. Figure from Mastrandrea *et al.* (2010).

In this final section of the Handbook, we look beyond the Best Practice Guide and consider the following examples which illustrate how to better cross-compare different studies:

4.1 Meta-analyses

4.2 Common garden experiments

4.3 Scientific community studies

4.4 The GAME project – examples of identical studies conducted in different locales

4.5 Local adaptation studies

4.1 META-ANALYSES

Successfully designing and completing single driver (i.e., performance curves to investigate modes of action, see Figure 2) and/or multiple driver experiment represents a tangible step forward towards an improved understanding of ocean biota responses to environmental change. However, this is just the initial step in the process of incorporating new findings into an expanded framework of knowledge. Before a consensus viewpoint can be reached by the marine global change community, further validation will be needed. This additional evidence could range from independent replication of the same experiment by other investigators, to supporting findings obtained using completely different experimental designs or methodology. At some point in this iterative process, novel findings may move from untested hypothesis to accepted theory. They may end up incorporated into broad, community-based data syntheses, such as the IPCC reports and other global or ecosystem-level assessments.

One way in which stand-alone results can get incorporated into a larger picture is through meta-analyses. These compilations of results from many published experiments have been used to discern statistical trends in the responses of marine organisms to single drivers like ocean acidification (Kroeker *et al.* 2010, 2013) and temperature (Thomas *et al.* 2012). Application of meta-analytical methods to multiple driver experiments is more challenging (Harvey *et al.* 2013), particularly as combinations, levels and ranges of drivers will vary between experiments, depending on the interests and inclinations of individual investigators. Also, there is the issue of spatio-temporal variability (for example see Vargas *et al.*, 2017) which reveals that there no sole control treatment for temperature or pH or other drivers. Thus, would-be authors of meta-analyses may be confronted with an 'apples and oranges' problem. Meta-analyses are also obviously most appropriate for relatively mature fields, with lots of published experiments- otherwise, major gaps may exist in the availability of data sets.

Meta-analysis is defined as:

> *"Meta-analysis is the quantitative, scientific synthesis of research results. Since the term and modern approaches to research synthesis were first introduced in the 1970s, meta-analysis has had a revolutionary effect in many scientific fields, helping to establish evidence-based practice and to resolve seemingly contradictory research outcomes."*
>
> From Gurivitch *et al.* (2018)

The following schematic provides insights into how meta-analyses are developed.

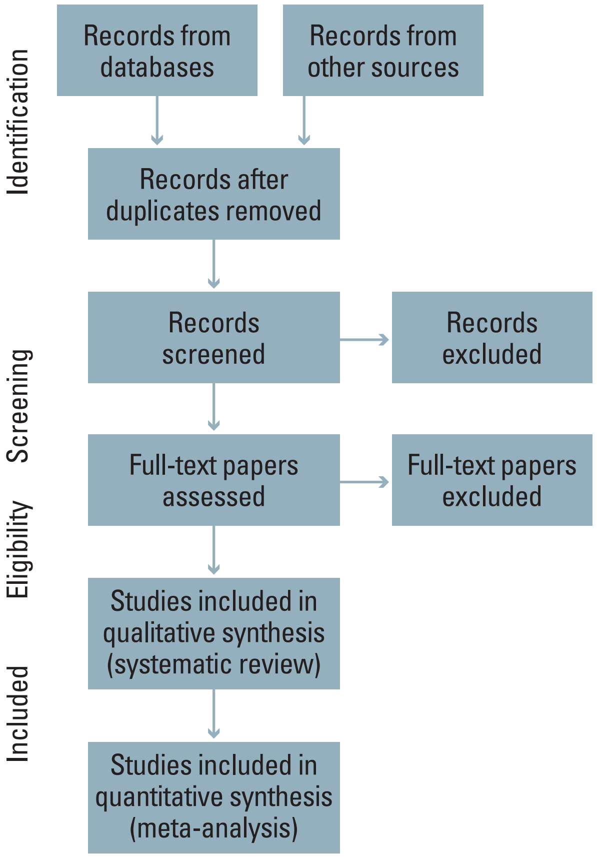


FIGURE 27. A PRISMA flow diagram, which describes information flow (the number of relevant publications) at the four stages of the systematic review process ('identification', 'screening', 'eligibility' and 'included').' Caption from Gurivitch *et al.* (2018). Flow diagram from Moher, D. *et al.* (2009).

For more information on meta-analysis see the recent review from Gurivitch *et al.* (2018).

4.2 COMMON GARDEN EXPERIMENTS

Another very useful way to compare findings within a single experiment (or set of experiments) is the common-garden approach. In this type of experiment, responses of a set of species or strains are compared under the same experimental conditions- hence the 'common garden' (Clausen *et al*. 1948). Common-garden experiments can be used to test whether a particular trait is determined by genetics or by physiological plasticity. They are also especially amenable to making well-supported generalizations about trends within and between taxonomic groupings. Examples in the context of ocean environmental change include comparisons of multiple sympatric (i.e., occurring within the same or overlapping geographical areas) species of phytoplankton from the Southern Ocean to iron and warming interactions (Fig. 28); the growth of four different species of dinoflagellates under the same CO_2 conditions (Fig. 29); and the nitrogen fixation rates of four strains of the cyanobacterium *Trichodesmium* across the same range of CO_2 concentrations (Fig. 30).

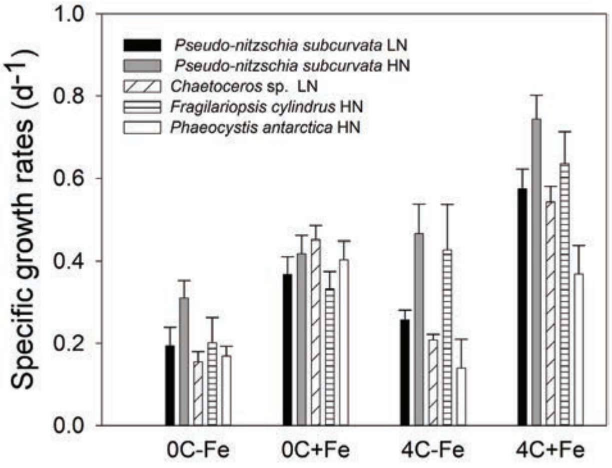


FIGURE 28. Specific growth rates of Antarctic phytoplankton in an iron/warming 'common garden' experiment, including: *Pseudo-nitzschia subcurvata* LN, *Pseudo-nitzschia subcurvata* HN, *Chaetoceros* sp., *Fragilariopsis cylindrus*, and *Phaeocystis antarctica* at 0°C–Fe limited (0C-Fe), 0°C–Fe replete (0C+Fe), 4°C–Fe limited (4C-Fe), 4°C–Fe replete (4C+Fe). Figure from Zhu *et al*., (2016).

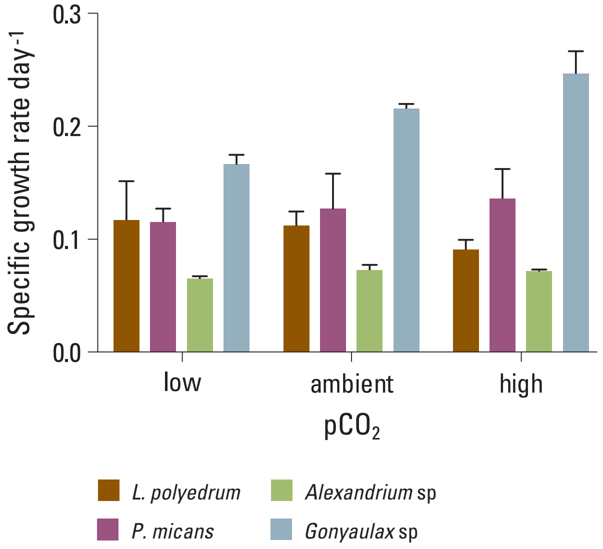


FIGURE 29. A long-term common garden experiment showing specific growth rates of four dinoflagellate species isolated from the same coastal bloom, following 8 months of selection at three CO_2 levels. From Tatters *et al.* (2013).

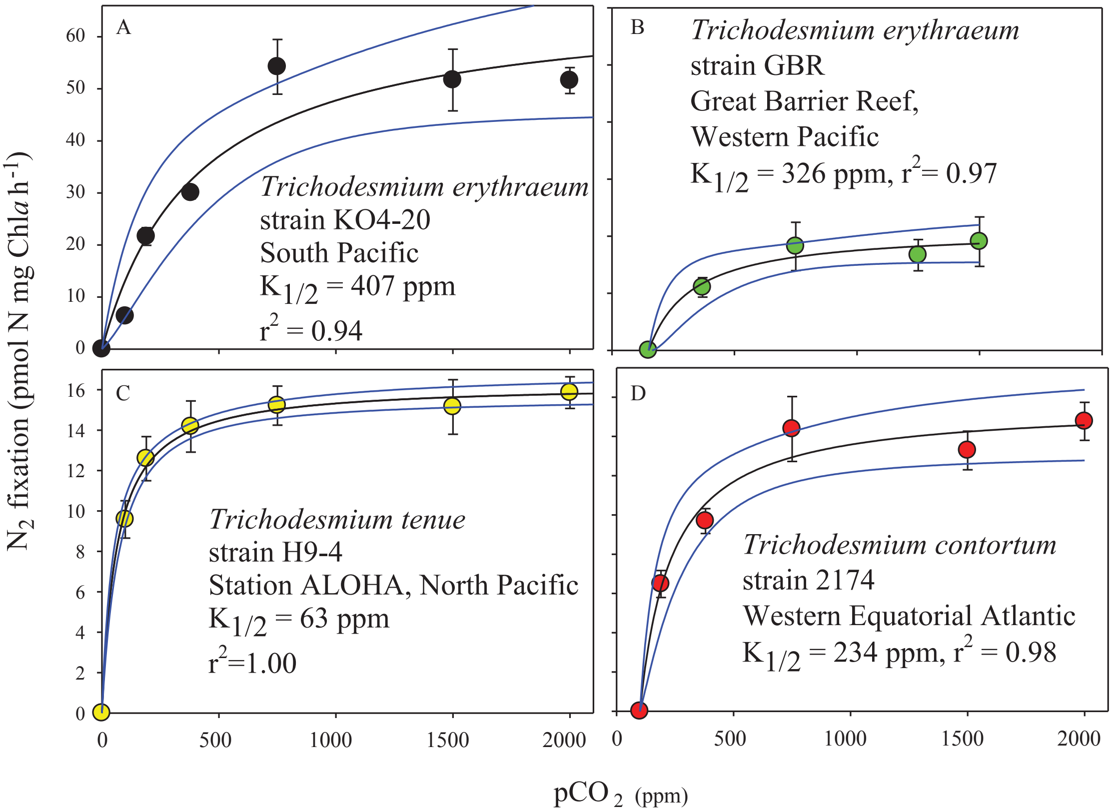


FIGURE 30. Nitrogen fixation rate response curves as a function of CO_2 for four isolates of the cyanobacterium *Trichodesmium* in a common garden experiment. Best-fit hyperbolic saturation curves (solid lines) with 95% confidence limits (dashed lines) for: *Trichodesmium erythraeum* a) KO4-20 and b) GBR; c) *T. contortum* 2174; d) *T. thiebautii* H9-4;. From Hutchins *et al.* (2013).

4.3 SCIENTIFIC COMMUNITY STUDIES

Other approaches to synthesize a large body of results is using a scientific community-based approach. One type of community experiment can be thought of as a geographically-dispersed common-garden experiment. Pooling the resources of numerous labs to do the same experiment with different isolates or taxa allows comparisons on a scale that is usually beyond what a single research group can accomplish. An example of this approach is presented in Boyd *et al*. (2013), in which an international set of researchers working at 9 different institutions obtained thermal response curves for 25 different eukaryotic and prokaryotic phytoplankton using common protocols. While care needs to be taken to ensure that experimental conditions are truly inter-comparable, the community experiment has the advantage of rapidly providing a large data set amenable to synthesis and statistical testing. Other approaches, discussed by Boyd (2013), include the use of biological reference organisms, standardised apparatus (such as FOCE, Free Ocean Carbon-Dioxide Enrichment, see Gattuso *et al*., 2014) or natural community large volume mesocosms.

4.4 **GAME** – AN EXAMPLE OF PARALLEL EXPERIMENTS CONDUCTED IN DIFFERENT LOCALES

Another approach to assessing how robust the findings of an experiment are is to repeat the same experimental design at different marine sites, where background conditions may vary subtly or in a pronounced way. This approach has been championed by the GAME programme which focuses on the coastal regions of our planet.

From their website.

> *"Several projects, for example, have studied factors influencing their biodiversity. GAME is also focusing on issues in invasion ecology and studies the ways in which environmental changes affect the interaction between species."*

https://www.geomar.de/en/research/fb3/fb3-eoe/fb3-eoe-b/game/game-about-game/

> *"GAME combines applied research with the academic training of young scientists. Every year, parallel research projects on current ecological issues are organised at different locations around the world. The research work is carried out by students who work together in bi-national pairs and who are supervised by scientists from GAME's partner institutes. The unique GAME approach provides generalizable insights into urgent ecological issues. GAME currently cooperates with 35 marine research institutes in 26 countries."*

4.5 LOCAL ADAPTATION STUDIES

Recent studies, such as Vargas *et al*. (2017) at a series of sites along the western seaboard of Chile spanning >20 degrees of latitude have explored how a range of species, locally adapted to natural site-specific variability, to ocean acidification. Such spatio-temporal variability revels that there is no sole control for pCO_2 but rather a suite of relative controls (Figure 31). A key message from the Vargas *et al*. (2017) study is that you must make the connection between the niche of an organism and it's environmental sensitivity to both single and multiple drivers since combinations of environmental conditions differ between regions (see Figure 1) but also populations which will be locally adapted to their matric of conditions. Thus, because of the role of local adaptation, it is problematic to extrapolate from one population to another. This issue must be tackled by comparing different populations of the same species (see Vargas *et al*., 2017 for details).

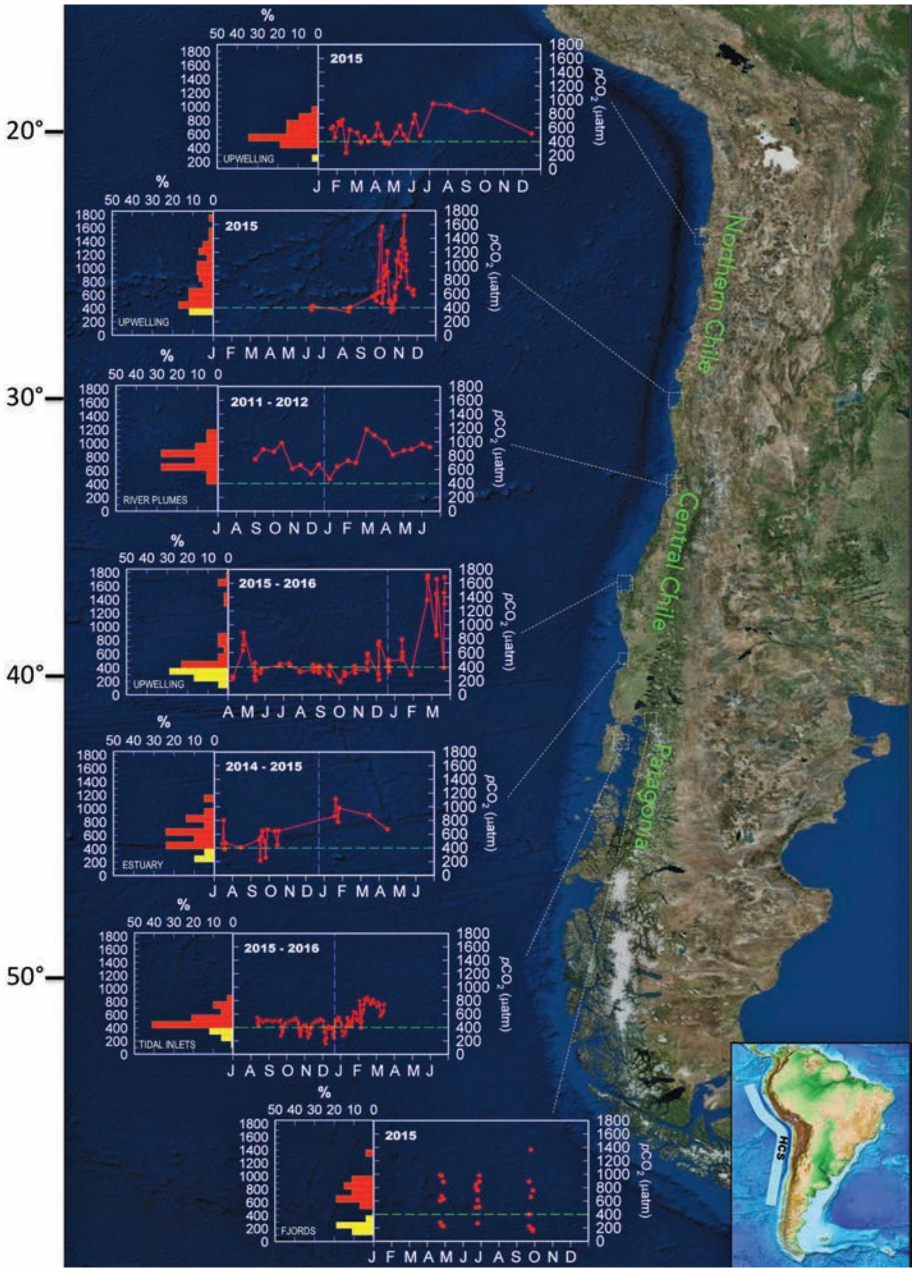
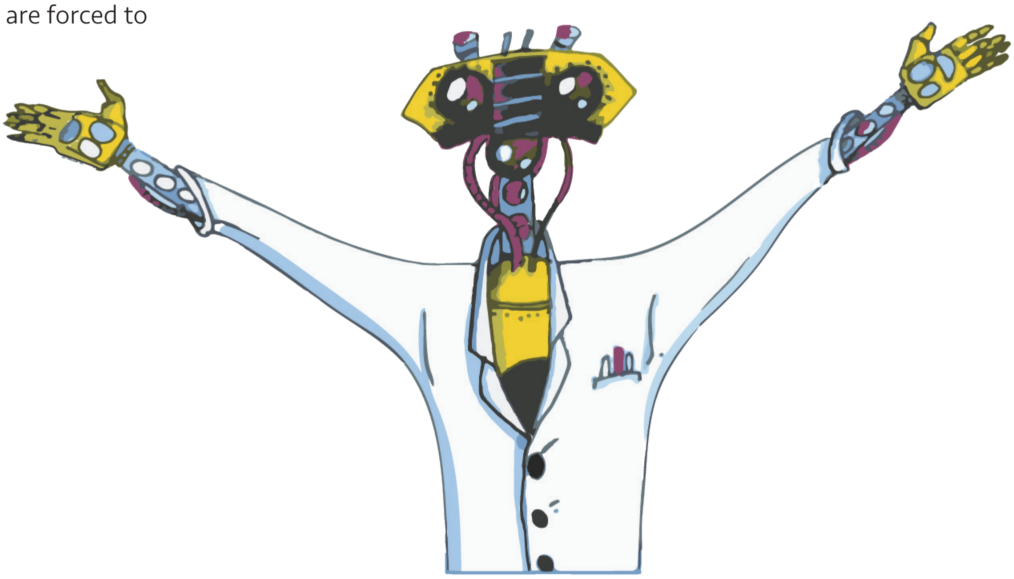


FIGURE 31. Temporal series (line plots) and frequency analysis (bars plots) of surface (upper 10 m depth) pCO_2 (µatm) for different coastal environments along the Chilean coast. Analysis was based on research cruises, field-monitoring programmes and buoys deployed in different coastal stations. The green dashed line in the temporal series represents the pCO_2 level of 400 µatm, the baseline level used as a control in most ocean acidification experiments. Dashed blue vertical lines represent the end of the respective year. Yellow bars in the frequency analysis correspond to frequency ranges < 400 µatm. Red bars highlight those pCO_2 frequency ranges higher than 400 µatm. Letters along the x axis represent months from January to December. Base map from Trackline Geophysical Data, National Centers for Environmental Information, NOAA https://maps.ngdc.noaa.gov/viewers/geophysics (From Vargas *et al*. 2017).

4.6 SUMMARY

Ultimately, our goal should be to design all of our experiments to help develop insightful new theoretical frameworks that increase confidence in the predictability of ocean ecosystem changes. If experimental results are framed and interpreted correctly, they can also be incorporated into increasingly realistic computer models (Boyd *et al*. 2008, Hutchins *et al*. 2013, Jiang *et al*. 2018). Such quantitative models can potentially be especially useful tools for framing policy responses and adaptation plans. Clearly, this is a strategy that will be needed increasingly in the near future as our human societies are forced to accommodate to the rapid changing ocean that we ourselves have created.



REFERENCES

Bopp L. *et al*. (2013) Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. Biogeosciences, 10, 6225–6245, 2013 www.biogeosciences.net/10/6225/2013/doi:10.5194/bg-10-6225-2013

Boyd, P. W. (2013). Framing biological responses to a changing ocean. Nature Climate Change, 3, 530–533. https://doi.org/10.1038/nclimate 1881

Boyd P.W., Rynearson TA, Armstrong EA, Fu F, Hayashi K, *et al*. (2013) Marine Phytoplankton Temperature versus Growth Responses from Polar to Tropical Waters – Outcome of a Scientific Community-Wide Study. PLoS ONE, 8 e63091. doi:10.1371/journal.pone.0063091.

Boyd, P. W., Dillingham, P. W., McGraw, C. M., Armstrong, E. A., Cornwall, C. E., Feng, Y. y., . . . Nunn, B. L. (2015). Physiological responses of a Southern Ocean diatom to complex future ocean conditions. Nature Climate Change, 6, 207–213. https://doi.org/10.1038/NCLI MATE2811

Boyd, P.W., S. Collins, S. Dupont, K. Fabricius, J.P. Gattuso, J. Havenhand, D.A. Hutchins, U. Riebesell, *et al*. 2018. Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change—A review. Global change biology, 24: 2239-2261.

Boyd P.W., T.A. Rynearson, E.A. Armstrong, F-X. Fu, K. Hayashi, Z. Hu, D.A. Hutchins, R.M. Kudela, E. Litchman, M. R. Mulholland, U. Passow, R.F. Strzepek, K.A. Whittaker, E. Yu and M.K. Thomas. (2013). Marine phytoplankton temperature versus growth responses from polar to tropical waters – Outcome of a scientific community-wide study. PLoS ONE 8: e63091

Boyd PW, Doney SC, Strzepek R, Dusenberry J, Lindsay K, Fung I (2008) Climate mediated changes to mixed-layer properties in the Southern Ocean: assessing the phytoplankton response. Biogeosciences, 5, 847–864.

Brennan G, Collins S. 2015. Growth responses of a green alga to multiple environmental drivers. Nature Climate Change. 5(9):892–97

Clausen, J., Keck, D.D. & Heisey, W.M. (1948). Experimental studies on the nature of species. III. Environmental responses of climatic races of Achillea. In: Carnegie Institute Washington publ. no. 581.

Cottingham, K.L., J.T. Lennon, and B.L. Brown. 2005. Knowing when to draw the line: designing more informative ecological experiments. Frontiers in Ecology, 3: 145-152.

Duarte C. (2013) In: The Conversation, http://theconversation.com/auditing-the-seven-plagues-of-coastal-ecosystems-13637).

Gattuso, J.-P., Kirkwood, W., Barry, J. P., Cox, E., Gazeau, F., Hansson, L., . . . Brewer, P. (2014). Free-ocean CO_2 enrichment (FOCE) systems: Present status and future developments. Biogeosciences, 11, 4057–4075. https://doi.org/10.5194/bg-11-4057-2014

Griffen BD, Belgrad BA, Cannizzo ZJ, Knotts ER, Hancock ER (2016) Rethinking our approach to multiple stressor studies in marine environments. Marine Ecology Progress Series 543:273-281.

Gunst, R. F., & Mason, R. L. (2009). Fractional factorial design. Wiley Interdisciplinary Reviews: Computational Statistics, 1, 234–244. https://doi.org/10.1002/wics.27

Gurivitch J., Koricheva J, Nakagawa S, Stewart G. (2018). Meta-analysis and the science of research synthesis. Nature, 555, 175-182. doi: 10.1038/nature25753.

Harvey, B. P., Gwynn-Jones, D., & Moore, P. J. (2013). Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. Ecology and Evolution, 3, 1016–1030. https://doi.org/10.1002/ece3.516

Hoar T. and D. Nychka (2008) – see https://gisclimatechange.ucar.edu/sites/default/files/users/Downscaling.pdf

Hönisch B, Ridgwell A, Schmidt DN, Thomas E, Gibbs SJ, Sluijs A, Zeebe R, Kump L, Martindale RC, Greene SE, Kiessling W, Ries J, Zachos JC, Royer DL, Barker S, Marchitto TM Jr, Moyer R, Pelejero C, Ziveri P, Foster GL, Williams B. (2012) The geological record of ocean acidification. Science,335(6072):1058-63. doi: 10.1126/science.1208277.

Hutchins, D.A., Fu F.X., Webb E.A., Walworth N., and Tagliabue, A. (2013). Taxon-specific responses of marine nitrogen fixers to elevated carbon dioxide concentrations. Nature Geoscience 6(9): 790-795. doi: 10.1038/ngeo1858

Jiang, H.-B., Fu, F.-X. Rivero-Calle, S., Levine, N., Sañudo-Wilhelmy, S.A., Qu, P.-P., Wang, X.-W., Pinedo Gonzalez, P., Zhu, Z., and Hutchins, D.A. (2018). Ocean warming alleviates iron limitation of marine nitrogen fixation. Nature Climate Change / 8: 709–712. doi.org/10.1038/s41558-018-0216-8.

Kroeker, K. J., Kordas, R. L., Crim, R. N., & Singh, G. G. (2010). Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. Ecology Letters, 2010(13), 1419–1434. https://doi. org/10.1111/j.1461-0248.2010.01518

Kroeker, K.J., R.L. Kordas, R. Crim, I.E. Hendriks, L. Ramajo, G.S. Singh, C.M. Duarte, and J.P. Gattuso. 2013. Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming. Global Change Biology, 19: 1884-1896. https://doi.org/10.1111/gcb.12179

Logan, M. 2010. Biostatistical design & analysis using R: a practical guide. London: Wiley-Blackwell.

Mastrandrea, M.D., C.B. Field, T.F. Stocker, O. Edenhofer, K.L. Ebi, D.J. Frame, H. Held, E. Kriegler, K.J. Mach, P.R. Matschoss, G.-K. Plattner, G.W. Yohe, and F.W. Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at <http://www.ipcc.ch>.

Moher, D. *et al*. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6, e1000097.

Mongin, M., Mark E. Baird, Bronte Tilbrook, Richard J. Matear, Andrew Lenton, Mike Herzfeld, Karen Wild-Allen, Jenny Skerratt, Nugzar Margvelashvili, Barbara J. Robson, Carlos M. Duarte, Malin S. M. Gustafsson, Peter J. Ralph & Andrew D. L. Steven (2016) The exposure of the Great Barrier Reef to ocean acidification. Nature Communications, 7,: 10732.

Quinn, G. and M. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press.

Riebesell U., Fabry V. J., Hansson L. & Gattuso J.-P. (eds) (2011) Guide to best practices for ocean acidification research and data reporting. [reprinted edition including erratum]. Luxembourg, Publications Office of the European Union, 258pp. (EUR 24872 EN). DOI 10.2777/66906

Riebesell U. and J-P. Gatusso (2015) Lessons learned from ocean acidification research. Nature Climate Change, 5, 12–14.

Sun, C., Feng, M., Matear, R. J., Chamberlain, M. A., Craig, P., Ridgway, K. R., & Schiller, A. (2012). Marine downscaling of a future climate scenario for Australian boundary currents. Journal of Climate, 25(8), 2947-2962.

Tatters, A.O., Schnetzer, A., Fu, F.-X., Lie, A.Y.A., Caron, D.A. and Hutchins, D.A. (2013) Short- versus long-term responses to changing CO_2 in a coastal dinoflagellate bloom: Implications for interspecific competitive interactions and community structure. Evolution 67-7: 1879–1891: doi:10.1111/evo.12029

Thomas MK, Kremer CT, Klausmeier CA & Litchman E (2012) A global pattern of thermal adaptation in marine phytoplankton. Science 338: 1085.

Thomas M.K. *et al.* (2017) Temperature-nutrient interactions exacerbate sensitivity to warming in phytoplankton. Global Change Biology, 23, 3269-3280.

Uthicke, S., Fabricius, K., De'ath, G., Negri, A., Warne, M., Smith, R., Noonan, S., Johansson, C., Gorsuch, H. and Anthony, K. (2016) Multiple and cumulative impacts on the GBR: assessment of current status and development of improved approaches for management: Final Report Project 1.6. Report to the National Environmental Science Programme. Reef and Rainforest Research Centre Limited, Cairns (144pp.).

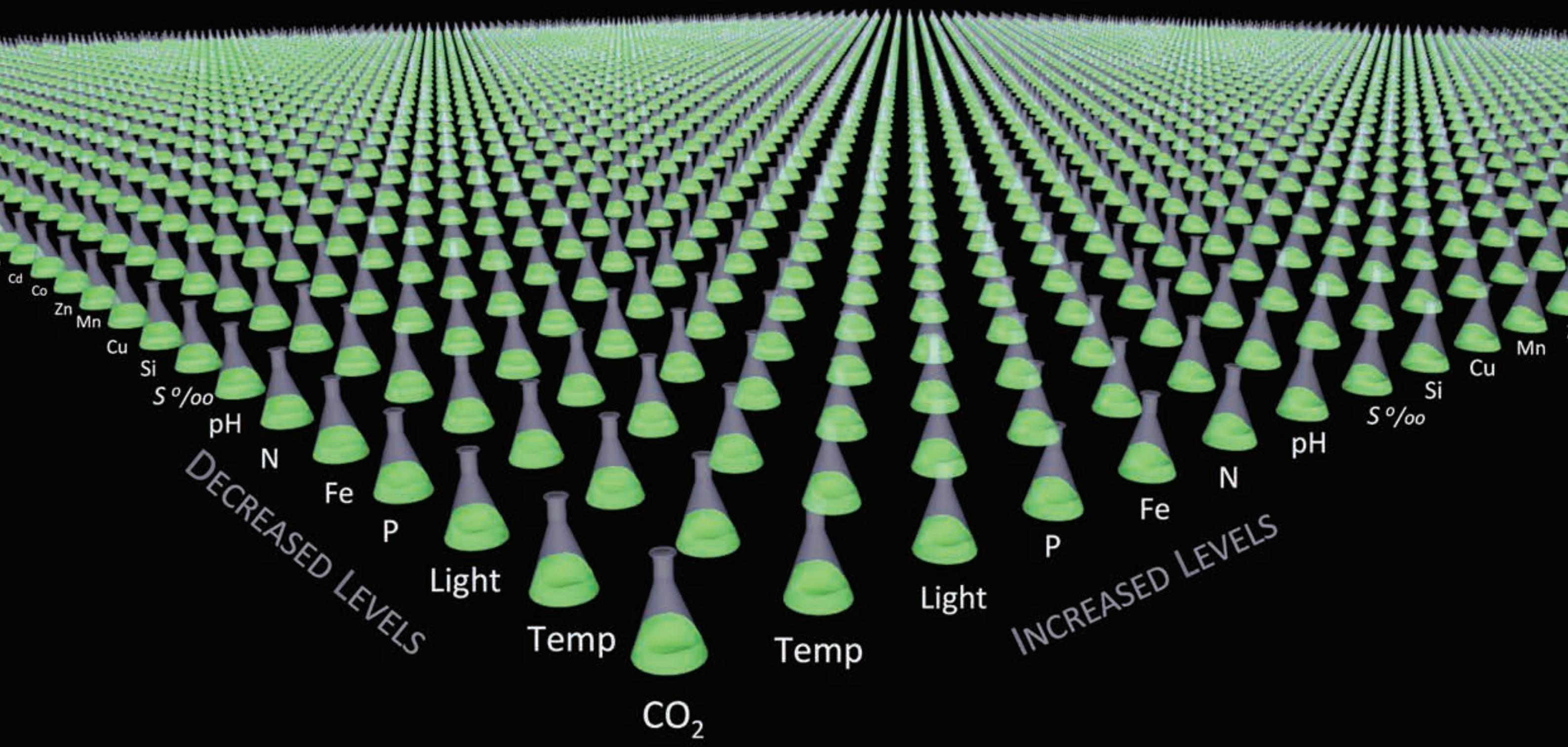
Vargas C.A., N.A. Lagos, M.A. Lardies, C. Duarte, P.H. Manríquez, V.M. Aguilera, B. Broitman, S. Widdicombe and S. Dupont(2017) Species-specific responses to ocean acidification should account for local adaptation and adaptive plasticity. Nature Ecology and Evolution, 1, 0084.

Underwood, A.J. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge, UK: CUP.

Zar, J.H. 2013. Biostatistical analysis: Pearson new international edition. Pearson Higher Ed.

Zhu, Z., Xu, K., Fu, F.-X., Spackeen, J., Bronk, D.A., Hutchins, D.A. (2016). A comparative study of iron and temperature interactive effects on diatoms and Phaeocystis antarctica from the Ross Sea, Antarctica. Marine Ecology Progress Series 550: 39-51. doi 10.3354/meps11732

Here’s to happy Meddling and to some great research based on your experimental designs!

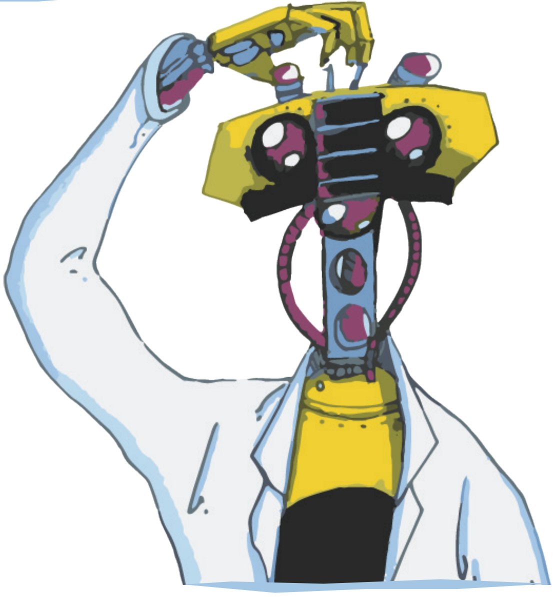


International Council for Science
SCOR
Scientific Committee on Oceanic Research

WG149

- Question(s)
- Inventory
- Design
- Experiment

- **Improved design**
- **Question addressed**



SCOR WG149
Handbook to support the SCOR

Best Practice Guide for Multiple Drivers Marine Research

March 2019