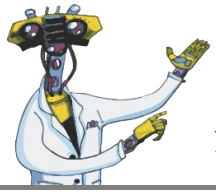


Video tutorials to support the

Best Practice Guide
for Multiple Drivers Marine Research

Data Analysis

Tutorial: The Data Analysis video tutorial can be found on the MEDDLE for Multiple Drivers Research YouTube channel.

Speaker: Peter Dillingham, University of Otago, New Zealand

Video: Christina McGraw, University of Otago, New Zealand

Transcripts: Rebecca Zitoun, University of Otago, New Zealand

Resources: The complete resources for the *Best Practice Guide for Multiple Drivers Marine Research* are available on the MEDDLE website.

0:00 - *Introduction*

By now you've watched Philip's and Jon's videos and you have got some idea about your experimental driver inventory as well as thought a bit about experimental design. Now we are going to dive into experimental design a little bit more deeply and think about why it matters.

A good experimental design will give you the ability to answer your question, maximise information, know and be able to defend your assumptions, reduce costs and more wisely allocate limited resources, get published, avoid catastrophe.

Text (0:20): A good experimental design will help you: answer your research question, maximise information, know and be able to defend your assumptions, reduce costs and wisely allocate limited resources, get published, avoid catastrophe.

0:41 – *Statistical terminology*

First, let's talk statistical terminology. A **FACTOR** is a variable that you plan on manipulating. In an experiment that examines changing ocean systems we often refer to factors as drivers. So in John's **3-FACTOR** experiment, he manipulated pCO_2, salinity and temperature.

Text (0:47): A FACTOR is a variable that you plan on manipulating. DRIVERS = FACTORS.

LEVELS are the values of a factor that will be applied during an experiment. For example pCO_2 might be set to past, present and future **LEVELS**.
Text (1:03): LEVELS are the values of a factor that will be applied during the experiment
TREATMENTS are the combination of factors at different levels that we actually apply during the experiment. For example, when Jon presented his major vector scenario, every culture flask represents a treatment.
Text (1:13): TREATMENTS are the combination of FACTORS at different LEVELS.

The number of **TREATMENTS** you have play a big part in how much effort or cost your experiment takes. So if you are going to do an experiment with a lot of treatments, you want to make sure that you have a good reason why.
Text (1:27): Generally, the more TREATMENTS, the greater the cost and effort.

What we want is an experiment that is as simple as it can be to answer the question we want. But no simpler.
Text (1:42): Design an experiment that is as simple as it can be. But no simpler.

An **EXPERIMENTAL UNIT** is the primary unit of analysis. This might not sound very well defined, because it isn't. You would be surprised how much time statisticians spend thinking about '*what exactly is 'n'?*' It's kind of why we are the life of the party.
Text (1:51): An EXPERIMENTAL UNIT is the primary unit of analysis.

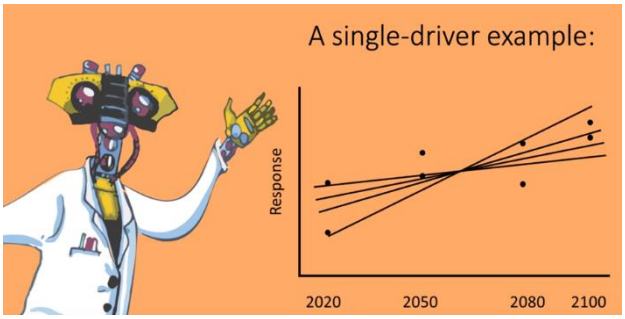
It also leads us to think about replication at different levels and the related issue of in-situ replication. Many global change studies involve replication at more than one level. For example, a culture tank study where you have a header tank controlling many culture tanks below it has two levels of replication. If this is ignored in the analysis, that is when in-situ replication comes in. If you want to find out more about in-situ replication, I refer you to the paper by Cornwall and Hurd in ICES journal of Marine Science.
Text (2:37): ICES Journal of Marine Science 73, 572-581 (2016)

2:46 – *Why do we care?*

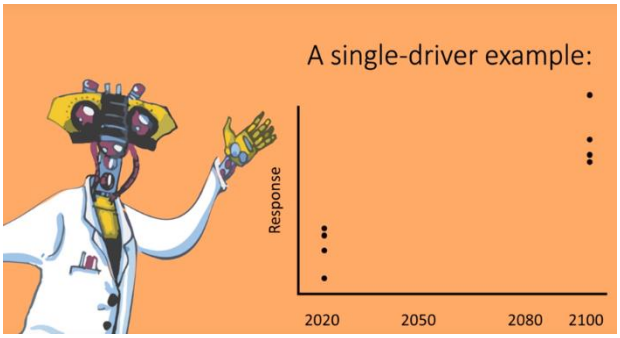
So why do we care about all of this? Well, we want to design an experiment that minimizes the effort and maximises the information we get. But we also need to be realistic about what we can archive.

Text (2:46): Why do we care? Our goal: minimise effort and maximise information

For example, you might want to do 4 levels of temperature, representative of let's say now, the future, and the year 2100, and a couple of points in between in order to get some kind of response norm. But you have a limited number of replicates, which might mean that your noise is overwhelming the signal and that difference across treatments. Well, we may be tempted to draw a curve that looks like this, but we could equally draw some other curves. Like these.



But let's say we cut it down to only two treatments, now and the year 2100. That gives us twice as many replicates in each treatment. By having more replicates, we may be able to get a useful answer, where before we didn't have enough information to say anything useful at all. So we need to be clear about how much noise we have as well as what kind of signal we are expecting.

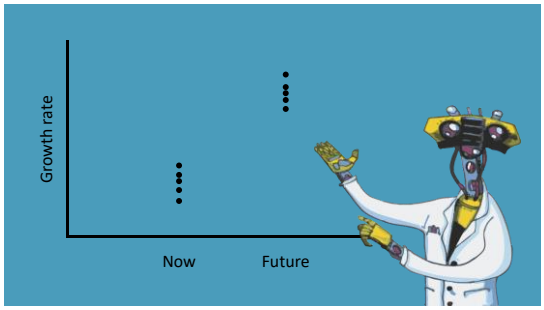


3:58 – *Design 1: The Scenario – 2 level scenario*

Let's take these ideas and jump right into one of the experiments Jon talked about, 'The Scenario'. This (*the 2-level scenario*) is a really good design if we want to just know what is happening between now and the future or two points in time.

Text (3:58): Design 1: The Scenario – Jon discussed the 5-level scenario in the Experimental Design tutorial. We will start with the 2-level scenario.

It's efficient, but remember we are getting very little information outside of those two points. So we won't get a response norm or any detail like that, but we will be able to say what is going on now versus the future. Because it [the 2-level scenario] is a nice simple set-up, this is also a good chance to talk a little bit more about some statistical details that are relevant in the other studies as well.



Let's start by generating some data. We will assume that we are looking at now versus the future and that we are measuring the growth rate of our organism. Let's say our experiments results look like this. Let's also assume that these points are independent. If they aren't, what we think is the now versus the future effect might just be a refrigerator by the window versus a refrigerator by the door effect.

How would we analyse this data? First, we need to remind ourselves of the original question.
Text (5:08): How would we analyse this data? First, remind yourself of the original question.

It might be '*Is there a difference between now and the future?*' Well, yes there is! If we are smart about what experiment we are looking at we are probably pretty sure that there is a difference between now and the future. The more interesting question is '*What is that difference*?' We have looked at a lot of previous work before we have started our experiment and so we have a good idea that there is going to be a difference.

Text (5:25): Instead of asking 'IS there a difference?' the more interesting question is 'WHAT is the difference?'

Hypothesis testing framework, i.e. looking at p values, is probably not what we want to do. It is not very interesting. It might not even be appropriate. If we know beforehand that the null hypothesis couldn't be true, then testing it isn't a good idea.

Text (5:39): A hypothesis test is NOT appropriate if you already know your null hypothesis is false.

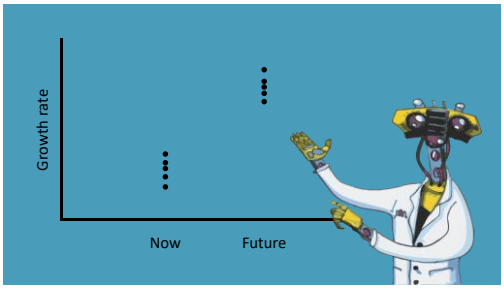
Text (5:53): Statisticians issue warning on p values – '…the p value …cannot determine whether a hypothesis is true or whether results are important' (Nature 531, 151 (2016)).

This leads us to something that statisticians are generally quite concerned with, which is an over reliance on *p* values.

Text (5:59): Moving to a world beyond 'p < 0.05' – '…it is time to stop using term statistically significant *entirely. Nor should variants such as significantly different, p < 0.05, and nonsignificant survive…" (The American Statistician 73, S1, 1-19 (2019))*

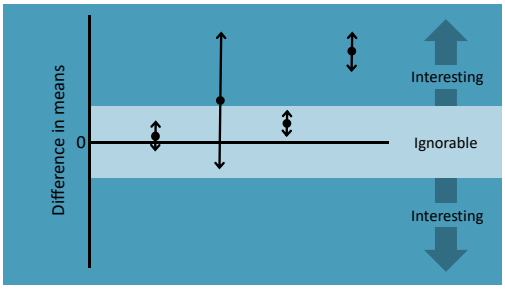
Instead we are much more interested in estimation and interpretation.

6:06 – *Estimation and interpretation*



So for comparing now versus future we can think about estimation and interpretation like this.

First we want to measure the difference in means of our two treatments. Zero is where there is no difference. We can also usually think of a value that, while different from zero, is ignorable from a biological point of view, but where values beyond that range start to become interesting. So let's look at 4 different point estimates and confidence intervals for that difference, which represent 4 possible experimental outcomes. Notice the first two [*results*] would return non-significant results, if we were to perform a null hypothesis test, while the second two [*results*] would return significant results.



Text (6:40): A hypothesis test tell us: (a) the two results on the left are not significant, 9b) the two results on the right are significant.

Text (6:59): The confidence intervals tell us much, much more.

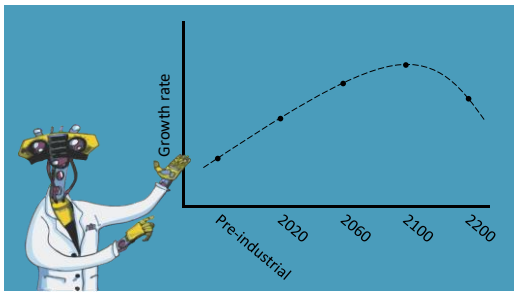
But the confidence intervals tell us much, much more. The first interval lets us know that the difference is positive, negative, or non-existent. But regardless, the effect is not important. The second interval lets us know that we probably didn't design our study very well. Not only do we not know if our result is positive or negative, we don't know if it [*the result*] is ignorable or important.

This is a good reason why you want to do power calculation before an experiment or pilot studies if you need to. But sometimes, even if you do those [*power calculations*], you can be unlucky. The initial estimates you put into your power calculation might have been wrong or you just happen to get a really weird data set.

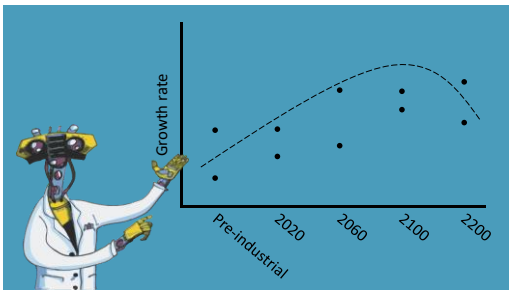
When we get to the third interval, we know there is a positive effect, but it is small enough that we have determined that it is not really that important. The fourth interval is the sort of thing that gets people excited. We have a positive effect and we know it [*the effect*] is of an important magnitude. So instead of simply doing a t-test and asking if there is a difference, let's look at a better question '*What is the difference between now and future?*', '*How am I going to measure it?*', '*Is that difference important?'*

Text (8:06): What is the difference between now and future? How will I measure it? Is the difference important?

8:06 – *Analysing data*



Let's make this scenario a little more exciting. By adding three more treatments to our scenario experiment, we have a chance of seeing the shape of the response curve.



However, biology is messy and depending on how much variability we have and how many replicates we have, we may or may not be able to see that shape. This is where the MEDDLE simulator can really help. Have a play with it and see if the data you might get is going to help you answer the questions you want.

You may have noted the extreme values we chose for our 5-point scenario are outside the conditions we are most interested in, now and 2100. This allows us to test extremes, providing more statistical power over all. Particularly, you may be looking for some kind of tipping point, local maxima or minima that may happen well into the future.

Analysing this [5-scenario] data may be a bit more complex than our 2-scenario approach. Once you collect your data you have a few options. You can simply analyse each treatment level independently as you would in an ANOVA. Or, you might want to decide to fit the response to a curve, maybe a straight line, maybe a quadratic, or some kind of smoother (*fit*) like a GAM, to describe your gradient.
Text (9:13): Analysing your 5-scenario approach: Analyse each treatment independently (e.g. ANOVA); Fit the response to a curve to describe a gradient, e.g. linear, quadratic, GAM.

Your analysis approach will depend on the questions you want to answer. It will also determine how many replicates you need and what sorts of assumptions you are making. Again, this is where the MEDDLE simulator can really come in handy: You can try different ideas, generate some data, see if you can analyse that data, and see if it answers the questions you are trying to answer. Remember, a little bit of time spent simulating can save you a lot of time in the lab. *Text (9:58): Simulation is the easiest way to save yourself time in the lab.*

10:05 – *Full factorial*

And now for something completely different. At the other end of the complexity scale we have got the full factorial design. Full factorials are good for examining all details, when resources are unlimited. That is, you can look at main effects, two way interactions, three way interactions, and so on. But they get ridiculously complicated very quickly.

If you have a full factorial, you can still do projections for now versus the future, however, your projections are no more accurate than if you just did a scenario with two levels: now and the future.
Text (10:29): Full factorial – You can still do projections for now vs. the future.

So if you think about that, what that means is, say you have got 3 factors, 2 levels for each factor = 8 treatments. And if you have got 5 replicates per treatment you have got a total of 40 replicates needed for a full factorial.

Text (10:47): If you have a full factorial with 3 factors and 2 levels per factor you have 8 treatments. With 5 replicated per treatment, you'll need 40 replicates for a full factorial.

If you are doing the 2-scenario you'll get the same level of precision with only 10 replicates.
Text (10:59): Only interested in now vs. future (i.e. 2-scenario)? Get the same level of precision with only 10 replicates.

11:02 – *Collapsed factorial*

Now, let's talk about the collapsed factorial. The collapsed factorial is a good design when you have one factor that you are most interested in, say a dominant factor, and a number of other factors that you collapse all together into one combined factor.
Text (11:06): Collapsed factorial: keep one dominant factor by itself and collapse all other factors into one combined factor.

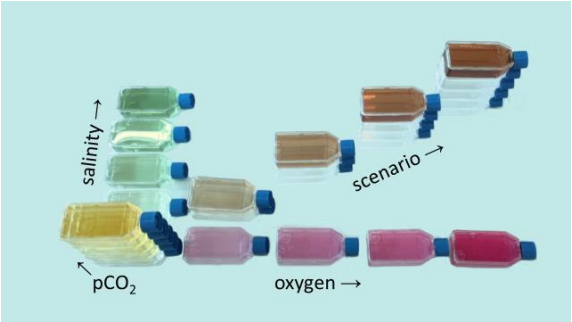
Collapsed factorials are good for projections and also give you some mechanistic understanding.
Text (11:20): Collapsed factorials are good for projections and also give you some mechanistic understanding.

For example, it can provide information about whether the putative dominant factor is actually dominant, or if the combined effect of other factors and their interactions with each other are important, some information about the interaction between the dominant factor and other factors.
Text (11:25): Collapsed factorials can provide information about whether the putative dominant factor is actually dominant. Collapsed factorials can provide information about whether the combined effect of other factors is important. Collapsed factorials can provide some information about the interaction between the dominant factor and other factors.

This is a good design if you have a limited number of experimental units, but you still want to get some separation between the dominant factor and all of the other factors.
Text (11:44): Collapsed factorials are good if you have a limited number of experimental units, but you still want to get some separation between the effects of the dominant factor and all of the other factors.

You have to accept that the other factors will all be confounded with each other. So the information that you will get is a little bit limited, however, you may have some other tools up your sleeve like proteomics that can help tease apart some of those other effects.

12:10 – *Major Vectors*



Next, let's talk about major vectors. Major vectors allow you to look at the individual factors as well as their combined effect. You would likely analyse each of the different vectors separately. So for Jon's example, we would look at the following: each of the 3 main effects (*pCO_2, oxygen, and salinity*) to get individual response norms and then we would look at the diagonal to get the scenario design.

12:10 – Experimental Design (or spend wisely)

One of the main take home points is that there is no single right answer for what experimental design to use.

Text (12:44): There is no single right answer for what experimental design to use

You want to think about what questions you really want to answer and which design is going to work best for you within the resources you have. For example, as we can see in the table, if you want to look at a 3-factor full factorial study and you need 5 replicates to have enough power, even 2 treatment levels per driver requires 40 experimental units. For a similar number of experimental units you could consider a scenario with 8 treatment levels and 5 replicates, or a collapsed factorial with 12 replicates and 12 treatment levels.

Text (12:54): For 3-factor study with 5 replicates per treatment: Alternatively, a collapsed factorial with 12 replicates and 12 treatment levels requires only 48 experimental units.

Levels per driver	Total number of experimental units			
	Full Factorial	Major vectors	Scenario	Collapsed factorial
2	40	25	10	20
4	320	65	20	20
6	1080	105	30	20
8	2560	145	40	20
10	5000	185	50	20
12	8640	225	60	20

Another way to look at it is that you may for example only have 24 culture tanks. So unless you replicate in time you will be constrained by the approaches available to you.

Text (13:20): Alternatively, if you have only 24 culture flasks, you can compare possible approaches.

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	8
		6	432	42	12	8
		8	1024	58	16	8
		12	3456	90	24	8
3	3	2	24	15	6	12
		4	192	39	12	12
		6	648	63	18	12
		8	1536	87	24	12
		10	3000	111	30	12
		12	5184	135	36	12
3	4	2	32	20	8	16
		4	256	52	16	16
		6	864	84	24	16
		8	2048	116	32	16
		12	6912	180	48	16
3	6	2	48	30	12	24
		4	384	78	24	24
		6	1296	126	36	24
		8	3072	174	48	24
		12	10368	270	72	24

Remember you don't need to answer every interesting question in one experiment. It is much better to answer one question well than a bunch of questions badly.

13:43 – Tips and Trick – or how I spend my Friday nights

You are probably here because you have got a really interesting question you want to answer. Expect that it will be complex and difficult to answer. And it could be a huge time investment. So the more planning you do upfront the more likely you are going to be happy at the end. Things that I recommend are:

- Know how to analyse your data before you begin your experiment
- If you have enough information to perform a power calculation, you should do so.
- Simulate data and analyse your simulated data. Remember you real data is almost never going to be as good as your simulated data. So if you cannot analyse your simulated data to answer the question you want to answer, you will be in serious trouble when you get to the real data later.

- Use the results from your simulation as a chance to refine your question and redesign your experiment
- Small pilot studies are your friend. You'll learn how your data will behave. For example, should you be operating in a log scale or a linear scale. They (*pilot studies*) will also give you a better idea about the level of noise you can expect and how many replicates you will need.
- For designs that don't have many treatment levels it is great to have a few extra replicates. Sometimes strange things happen and you might lose a culture or you might actually find out something really interesting.
- If you have the need and the opportunity to consult with a statistician or a senior scientist who knows some statistics, then you want to do this before the experiment

Finally, be brave and don't be afraid to try something new. That's what is going to make your experiment really exciting to everyone else.