
THE BAREFOOT ECOLOGIST'S TOOLBOX

DR JEREMY PRINCE

LAST UPDATED : MARCH 2021

Contents

5	PREPARING THE DATA FOR LBSPR ASSESSMENT
7	Introduction
9	Maintaining Data Quality as You Go Along
11	Preparation for LBSPR Assessment.
19	Choosing the Best Estimates of Life History Ratios (LHR)
21	References



C H A P T E R S E V E N

PREPARING THE DATA FOR LBSPR ASSESSMENT

Introduction

This section is very applied, a practical work-horse chapter where the real barefoot ecologist starts working up and deriving meaning from the data set hard-won from the fishing. For the reader wanting to apply the spawning potential surveys methodology, the previous chapters should have enabled the collection of data for LBSPR assessment, this chapter should enable you to prepare the data you have collected into the inputs needed for LBSPR assessment. Depending on your perspective, this will either be the most immediately useful chapter, or the one without any theoretical interest.

Over time I have developed some data-handling and analytical tools in Excel to assist with this data preparation process. The aim of this chapter is to provide a manual to support using those tools. By the end of this chapter you should have all the inputs you need ready for use with the free web-based LBSPR assessment software described in the next chapter.

The chapter starts with discussion about the importance of working continually with your data gatherers to ensure the data being collected is of a useful high-quality. It highlights some of the common data quality issues you are likely to confront and describes how to correct them.

The chapter then plunges in a step-by-step instructions for using spreadsheet routines:

- to estimate size of maturity (L_m), and
- prepare data for presentation to the LBSPR assessment software.

I have tried to make this chapter the sort of things I would do in my training workshops, talking people through my own process step-by-step. The idea is that you should upload, the various Excel spreadsheets discussed, and as you read this chapter be working through them using your own data sets. The implicit but important assumption here, is that the reader will have used the data entry templates discussed in the previous chapter and downloaded from the 'biospherics.com.au website'. That should ensure that the data-base you are developing is in the format I am describing here, so that this step-by-step guide is pertinent. Although there is one thing to be remembered. I

Maintaining Data Quality as You Go Along

In the initializing training workshops there should have been discussion with community observers about the importance of the quality of their data recording, including how to measure fish to the nearest 1 mm.

Start off Working Alongside the Fish Measurers

Never-the-less, by themselves, those initial training workshops with communities will have provided too little practical experience to ensure good data quality. The organizer of the workshop should plan to work in the field alongside village data gatherers after the initial training, to follow-up, with some in the field support, supervision and correction. This could involve several days fishing together as a training team and then measuring and eating the sampled fish together. Or a field officer visiting each measurer in their location and spending a couple of days measuring and gauging the fish together. This will allow basic trouble shooting to be done for the measurers.

Interrogating the Data regularly

Subsequently, keep visiting the observers regularly to collect their data sheets and talk with them. Upon returning to your office enter all the data that has been collected into the database and start doing preliminary analyses with them right from the beginning. Data are like muscles, if you don't use them regularly, when you eventually get around to using them in the end, you will find they don't really work. So exercise your data regularly, interrogate them and see what they can tell you by practicing the analyses explained below:

- Use the internet to work out how to do a pivot-table in Excel and make one summarising how many of each species have been measured. What are the main species being measured? Look down the list of species names arranged by alphabetical order. Notice anything strange? Can you see slightly different spellings of the same species appearing as if they are multiple species? That will need correcting in the database.
- What sizes are the main species being measured? Using the directions provided below plot up the size-frequency histograms of the main species regularly. Are nice smooth bell-

work with a little Mac Air Book which is great to travel with, but consequently I use a version of Excel from Microsoft Works for Mac. So please don't be surprised if some of the detail in my instructions about what to do in Excel, do not exactly describe your PC machine and Excel running in its native Microsoft environment. You may need to think more laterally about what I am doing with Excel in Mac-land, and then figure out for yourself how to do the equivalent with your own software.

shapes emerging from the samples as the numbers mount, as expected? If not, why not? Are you seeing big spikes emerge at every 5 or 10mm of length?

- What do the estimates of size of maturity look like so far?

Keep asking yourself what makes sense about what you are seeing, and what doesn't? Doing these types of exercises regularly, and asking these types of questions of the data during the data collection program, will enable you to detect and correct errors early. If you do not, you will inevitably get to the end of the data collection program and then find all those little errors degrading your data and their usefulness.

Each time you visit your team of measurers take your preliminary results and observations and show them. Give the measurers plenty of feed-back about the data they are gathering. Let them know what you are seeing, and find out whether what you think you are seeing makes sense to them. They will have a keen sense of what they are measuring, and may well alert you to problems in your data-base. As the data accumulate, you should arrange more workshops with the community to work with, and review, the data. The community will enjoy watching the scientific process of the data slowly coming together, and in this way, by the time you finalise results, they will be more invested in them, having watched the process throughout.

Preparation for LBSPR assessment.

The aim of this section is to provide a step-by-step guide to analysing the length composition and size of maturity data collected for fish with a sheet in the Excel Workbook called '1. Excel template for fish data with analysis sheets' that can be downloaded from the 'Data Collection and Analytical Techniques' page of the 'biospherics.com.au' website. The sheet in the workbook used for this process is called 'Analyses'.

Making a Length Frequency Histogram

In this initial process the spreadsheet provided will be used to generate a length frequency histogram of your data to view, and prepare a column of data for later export into the LBSPR assessment.

1. Open the workbook called '1. Excel template for fish data with analysis sheets'.
2. In the Date Entry spreadsheet select and highlight all columns.
3. Go to the Data tab and pull-down, select sort, and then sort by Scientific Name / Maturity / Sex / Length.
4. In the same Data Entry spreadsheet find the species you are interested in and then go into the Analysis spreadsheet.
5. In the Analysis spreadsheet prepare the length sample by first selecting and highlighting cells B3:B52.
6. Enter the Equation Bar to modify the pre-existing equation for making a length frequency by clicking just to the left of the comma in the equation (inside the brackets) and then deleting back to left bracket. But do not delete the left-hand bracket.
7. Next go back into the Data Entry spreadsheet and select (highlight) all the length measurements in Column D for the species you have chosen to work with. NB: a common error because you are focussed on the species name is to select the species names and not their lengths.

8. Once you have selected all the lengths you want summarised in the length-frequency histogram, hold down Shift + Control, and at the same time press Enter.

9. You should then get taken back to your completed length frequency in the Analysis spreadsheet.

10. You should now be able to see the size composition plotted up in the lower of the two charts to the right of the Analyses spreadsheet. You may want to adjust the range of data being plotted by clicking on the chart and then using Chart Tools on the toolbar and the Select Data option (in Windows) or Chart and Source Data in Macintosh. Adjust the length of the plotted series as required.

Look for Spiky Length Composition Data

You can change the number and width of length bins used to make the length-frequency histogram by changing the value in cell A1, that will change the resolution of your histogram. Try a few different resolutions say; 1, 2, 5, 10mm. Watch out for the histogram becoming spikey at some resolution, as illustrated below in figure 1. This spikiness will almost inevitably indicate that your observers are getting complacent and, instead of really measuring each fish to the nearest mm, they have started rounding up the measurements, to the major marks on their measuring tape. You should be able to see the extent to which this is occurring by noting whether the spikes occur every 5, 10, 20, 50 mm. Almost inevitably observers round-up their measurements to some extent. The LBSPR algorithms cannot account for this human error and it will result in the accuracy of assessments being degraded. Ideally you want the observers to be measuring to the nearest 1mm, but in practice if the fish you are assessing are generally >400mm, rounding to the nearest 5mm will probably be acceptable. Note the finest resolution which gives you a relative smooth length-frequency histogram, because this is the resolution you are going to want to use for your LBSPR assessment.

If you are detecting rounding errors, go and hold more meetings with your observers to give them feedback, show them what is happening to the data, as well as example of what good

smooth length-frequency histograms should look like. Let them know you can see from their data what is going on, and that you care about their data enough to be taking note.

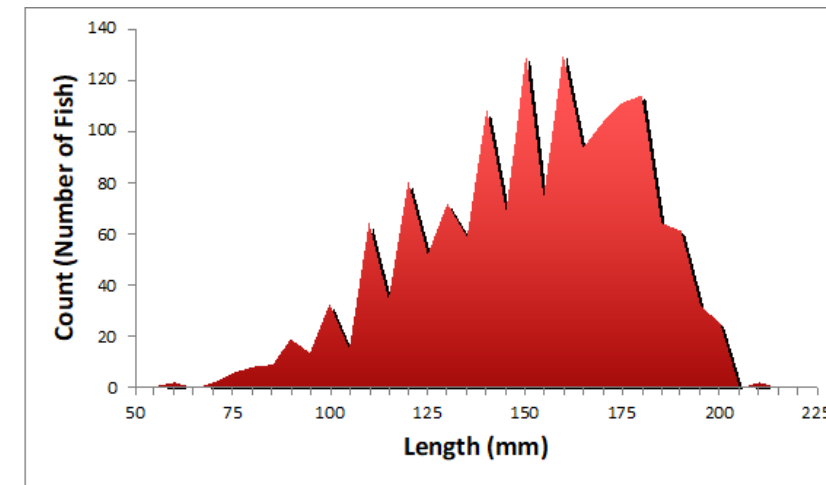


Figure 1. Data collected with strong regular spikes at 10mm intervals indicating the data collectors have tended to round measurements to the nearest cm.

Extracting Length Data for the LBSPR Software

Next extract the length frequency data and save it into the comma separated value format (.csv) that is needed for the LBSPR web-based software. Before you do these next steps, use cell A1 to set the bin width to the smallest width that gives you a smooth histogram. Ideally you should have >10 length categories or 'length bins' and your data should be a nice smooth uni-modal curve, try different bin widths and try to find the width that comes closest to the objective of maximizing the number of bins (i.e. minimizing bin width) and producing a smooth unimodal curve, and then:

11. Select and copy the values in cells A2:B52.

12. Open a new workbook and paste the copied values into the top left-hand corner of the new spreadsheet (into cell A1). NB: Use the Special Paste / Values Only option in excel. Do not copy and paste the underlying equations.

13. Save the new data file in comma delineated format (.csv). Note this file must be saved in .csv format and not excel format.

Alternatively, you could also have gone back to the Data Entry sheet and selected all the raw length measurements in Column D for the species you have chosen, and then copy and pasted them into the .csv data file you are creating (i.e. steps 12 & 13). The LBSPR software can use either type of data format, and if you do choose the list of length measurements format, you will retain flexibility with regard to the bin width used in the analysis.

Size of Maturity Analysis

In this next sequence of the data preparation process we use the same Analyses spreadsheet to analyse the macroscopic size of maturity data to:

- estimate proportion mature by size class, and
- fit the trend in proportion mature by size class to a logistic, or S-shaped, curve to estimate L_{50} and L_{95} .

The logistic size of maturity curve is fitted to the data by using a simple sum-of-squares minimization routine that was first taught to me by Ray Hilborn and Carl Walters. My aim here is to simply tell the reader how to use this spreadsheet, rather than to explain the detail of what each part of the analysis does. To some extent I am assuming readers wanting that level of detail will be able to reverse engineer the spreadsheet and look through the embedded equations to work that out for themselves. To help with that process I also provide for downloading another spreadsheet called '11. BSC L50 Estimation for Cyril' which I prepared for Cyril in the Philippines. With that workbook, the sequential spreadsheets show and explain column by column how the fitting routines are constructed.

Here we will just leap straight back into instructing the reader in how to do it.

14. In the Analyses spreadsheet start the size of maturity analysis by making length frequencies of Adult females (column D3:52), adult males (column E3:52) and immature (column F3:52).

15. For each of the three columns repeat steps 6-9 above taking care to select in sequence the lengths of adult females,

adult males and immature fish of the chosen species.

16. Examine the estimated trend in percentage mature by size in Column H and see what it tells you. Hopefully it shows a relatively smooth transition from 0% mature through to close to 100% mature. In some cases the underlying trend might be improved by removing (deleting) any clearly outlying estimates based on just a couple of data points. Often the smoothness of the trend will be improved by increasing the bin-width (Cell A1) so that each length class contains a larger sample.

17. Alternatively, no matter what you do, Column H may show that some high and relatively trendless proportion of all size classes are mature. Do not despair this is quite common with tropical reef fish, and indicates the fish in your sample only become vulnerable to being caught as they mature. In that case you may want to use the left-hand side of the length frequency curve to infer the proportion mature. For this you use the proportional length-frequency values in Column C, rather than percent mature data in Column H. This option, how to apply it, and the rationale for applying it are discussed in more detail in the next section.

18. Whichever option you eventually choose you will end up back here selecting and copying values from either Column C or H.

19. Then you will do a Special Paste / Values Only of the selected values into Column I which is labeled 'Selected Values' only.

20. Clean up the values in Column I by removing any old values left over from a previous analysis. You may also have ended up with a 'Division by zero' error message for some size classes without any data which will need deleting.

21. Having pasted your selected values into Column I you will notice Predicted Values in column (J). These are predicted for the same length classes in Column A using the logistic equation shown in Cell M1 and the parameter values shown in Cells L3, L6 & L9. The first of these parameters 'A' defines the maximum value the logistic curve will predict, in this case 100%

percent mature, and is left fixed at this value in this analysis. The second parameter 'r' defines the steepness of the logistic curve and the third parameter defines the size at the half way point of the logistic curve, in this case the size class in which 50% are predicted to be mature (i.e. L_{50} or L_m). You can give yourself a feel for how this equation works by trying out different values of these three parameters and watching how the plotted logistic curve (line) in the upper chart to the left changes with different values. The process of fitting described below is visualized by this plot which shows the selected data (Column I) with red squares, and the Predicted Values (Column J) with blue triangles connected by a curve.

22. In Column K the squared differences between matching Selected Values (Column I) and Predicted Values (Column J) are calculated, and the sum of all the squared differences is computed in Cell K1. Clean up Column K by making sure squared differences are only estimated for length bins where you actually have data.

23. Now start the fitting process with the aim of matching the predicted values in Column J as closely as possible to the selected values in Column I. This is done by testing different values of L6 & L9 to minimize the sum of all the square differences in cell K1.

24. First make sure the A value in L3 is set to the right value. If you are using the Percentage Mature data estimated in Column H, then the A value in L3 should mostly be set to 100%. But in some cases you might know that a large constant proportion of larger animals are inaccurately judged to be immature, so that percentage mature values top out below 100%. In that case you might choose to set and fit to a lower value for parameter A; say 75-95%.

25. Next start adjusting the slope (r value) in L6 and L_{50} value in L9 until the sum of the Squared Differences in Cell K1 is as small as it can possibly be made. This can be done manually by trial and error, adjusting first one value and then the other, again and again until you reach the lowest possible value of Cell K1. And I recommend you do it a few times manually like that while you get the idea of what the fitting process of minimizing a sum of squares is all about. In the long term you will probably

want to automate this process by using the Solver Function, which can be downloaded and added into Excel, and then will be found under the Data Tab.

26. Finally, by looking down Column J determine the size class at which 95% of the fish are predicted to be mature (L_{95}) and make a note of this. If you chose to fit the curve with $A=100\%$ the value you will be looking for in Column J will be as close to 95% as possible. You may improve your precision in this process by reducing the bin width in Cell A1 to smaller values. Note that if for some reason you happened to set $A < 100\%$ in Cell L3 then the value you are looking for at this stage will be $0.95 \times A$ and less than 95%.

Estimating Size of Maturity Direct from Length Frequency Data

Many Indo-Pacific reef fish species exhibit ontogenetic habitat shifts, and have juveniles that grow up in some shallow nursery habitat that provides the food resources and protection they need, such as shallow mangroves, seagrass beds or coral rubble flats. It is only as they mature that they then move out into the coral reef habitat where they become vulnerable to being caught by fishing (Nakamura et al. 2008, Grol et al. 2011). This type of behaviour can mean that juveniles are rarely caught and so difficult to sample. Which in turn makes it difficult, or impossible, to define the transition from 0 to 100% mature, preventing the estimation of L_{50} by examining gonads either macroscopically or microscopically (e.g. Williams et al. 2008; Currey et al. 2013; Moore et al. 2014; Taylor et al. 2018).

This is the reason that at Step 17 above it is sometimes impossible to discern any trend in the proportion mature; because being relatively mature was a pre-condition for being observed in your sample. Fortunately, in this situation the observation that only maturing animals occur in your sample, means that the size composition can be used to infer size of maturity. With a recently completed, but as yet unpublished study in Palau, along with colleagues Steven Lindfield and Bill Harford, we demonstrated that when you regard the left-hand side of the catch length frequency histogram, as a cumulative

Choosing the Best Estimates of Life History Ratios (LHR)

The final input you need to prepare for completing an LBSPR assessment are estimates of the life history ratios (LHR) of your species; M/K and of L_m/L_∞ . Together these values define the shape and size of a length frequency histogram and so the accuracy of LBSPR depends upon having accurate estimates for these two input parameters.

Naturally the reader is going to want to use some study of the chosen species to estimate these parameters, and will prefer a local study if it exists, perhaps even their own study. After all we fisheries biologists have all been trained up to believe that local species-specific studies are best, and to have unlimited confidence in our own ability to complete good studies. The sad reality is that a lot of the published estimates of growth and mortality on which these estimates are based, are not high quality. There are many challenges to producing high quality estimates of life history parameters and life history ratios, which most of us who began our careers doing these types of studies, were not fully aware of, and did not account for, when we made the studies. Length-based studies have implicit assumptions that bias resulting estimates, as does size selective sampling, sampling from fished down populations, or not accurately constraining the size at which growth begins. To mention just a few of the main reasons that degrade published growth estimates, and of course, all the growth parameters are correlated, so if one is inaccurately estimated, all the estimates end up being biased.

I know all this now, because I have spent the last six years collecting thousands of studies of fish age and growth, and where-ever possible re-digitizing their data, so that I could apply a standardized range of correction factors, and quality control criteria, to derive the best possible estimates of the LHR for species, genera and families. I was much less aware of the issues when I was publishing studies of age and growth, and my recent meta-analysis shows I was not alone. My advice is to be very careful about estimating the LHR of your own species from a small number of local studies, especially if they are length based studies of growth, based are on moderate to heavily fished populations, or worse borrowed from FishBase..

Far better to use estimates derived from a meta-analysis

frequency curve, the 50th percentile of that curve gives a good approximates of the size of maturity estimate you would derive (if you could) with histological techniques.

With the last few steps below, I explain how you can use the same 'Analyses' spreadsheet to do this for yourself if you do find that the macroscopic inspection of your species' gonads reveals a high and trendless proportion of each size class is mature.

27. In Column B find the largest frequency count, which will correspond to the modal size category of the length frequency (i.e. the top of the main mode of the length frequency histogram), and enter that value in Cell C1. This will normalize Column C into a frequency histogram with a peak value of 100, and you will be able to use the ascending values on the left-hand side of the main mode as a proxy for percent mature.

28. Copy the ascending values from 0 – 100 from Column C and Special Paste / Values Only them into the corresponding cells of Column I. And then resume the process from Step 18.

References

Currey, L.M., Williams, A.J., Mapstone, B.D., Davies, C.R., Carlos, G., Welch, D.J., Simpfendorfer, C.A., Ballagh, A.C., Penny, A.L., Grandcourt, E.M., Mapstone, A., Wiebken, A.S., Bean, K. 2013. Comparative biology of tropical *Lethrinus* species (Lethrinidae): challenges for multi-species management. *Journal of Fish Biology*, 82: 764-788.

Grol, M.G.G., Nagelkerken, I., Rypel, A.L., Layman, C.A. 2011. Simple ecological trade-offs give rise to emergent cross-ecosystem distributions of a coral reef fish. *Oecologia* 165:79-88.

Moore, B., Rechelluul, P., Victor, S., 2015. Creel survey and demographic assessments of coastal finfish fisheries of southern Palau, September 2014. Secretariat of the Pacific Community

Nakamura, Y., Horinouchi, M., Shibuno, T., Tanaka, Y., Miyajima, T., Koike, I., Kurokura, H., Sano, M. 2008. Evidence of ontogenetic migration from mangroves to coral reefs by black-tail snapper *Lutjanus fulvus*: stable isotope approach. *Marine Ecology Progress Series* 355: 257-266.

Taylor, B.M., Oyafuso, Z.S., Pardee, C.B., Ochavillo, D., Newman, S.J. 2018. Comparative demography of commercially-harvested snappers and an emperor from American Samoa. *PeerJ* 6:e5069; DOI 10.7717/peerj.5069

Williams, A.J., Currey, L.M., Begg, G.A., Murchie, C.D., Ballagh, A.C. 2008. Population biology of coral trout species in eastern Torres Strait: Implications for fishery management. *Continental Shelf Research* 28: 2129-2142.

of many high-quality studies, and my purpose for the meta-analysis just completed was to make high quality LHR estimates available for other people to use. Aiming to develop as many high-quality standardized estimates of the LHR as possible I collected >1335 published studies and applied rigorous standardization and quality control procedures to develop a database of 1576 and 861 high quality estimates of M/K and L_m/L_∞ , respectively. That study, which should be published in *Fish & Fisheries* during 2021, conclusively demonstrates that the two LHRs are strongly correlated and that they vary predictably by taxa, with the level of family having the most predictive power. The reader can access our high quality LHR estimates for 69 families of fish by downloading the Excel spreadsheet called '10. Best LHR estimates for fish Families' from the 'Data Collection and Analytical Techniques' page of the 'biospherics.com.au' website.

My next challenge is to start compiling an equivalent database for invertebrate taxa as well, but it will probably be some time before I have completed that task. In the meantime feel free to contact me and ask whether I have any LHR estimates for the invertebrate taxa you are working on.