

Fig. 2. Relationship between and index of gill surface area and an index of growth performance in scombrids (broken line) and in miscellaneous marine fishes (solid line). Points and computational methods are documented in Hughes, G.M. and M. Morgan (1973 Suppl. Publ. SUP90005, British Library, Lending Div., Boston Spa) and Pauly, D. (1981, Meeresforsch. 28(1):251-282). The *Latimeria chalumnae* outlier (Δ) is based on growth parameters estimated from size-at-age data in Hureau, J.C. and C. Ozouf (1977, *Cybium* 3(3):127-137) and gill area estimates in Hughes, G.M. (1972, *Experientia* 28:1301-1302). A correction of gill area estimate (Δ) for a typographical error (mm^2 to cm^2) renders ordinate value credible, but only in association with a much reduced estimate for the abscissa i.e., a much lower growth performance.

material from theses should be used only if the same material is not available in published version and the author can be contacted and his/her approval obtained, or if the thesis is more than 2-3 years old and can be expected to remain unpublished (as do the majority of theses, unfortunately).

- Written permission to cite certain sources is required (by some publishers) for preprints, such as the often much cited papers presented at the annual Council Meetings of the International Council for the Exploration of the Sea.

- Use "pers. comm." with name of author who provided you personally with a communication and check your wording with the source person. He or she may also be mentioned in the acknowledgement section of your paper.
- For any other unpublished data, it is most important to try hard to get explicit (verbal or written) permission to use the data.

The Error of Solipsism

Fear of (other people's) errors has been mentioned to me by colleagues (who evi-

dently trust themselves a lot) as the key reason why they would never use other people's data. This is a rather immodest attitude. For my part, I trust, for example, G.M. Hughes' ability to conduct and/or evaluate work on the gill area of fish more than mine. I relied on his data rather than spending years to derive a few more data points whose "fit" could not even be evaluated, given that Fig. 2 would not exist.

Besides, the whole attitude smacks of solipsism (which means to believe that we only imagine the world around us, as postulated by Bishop G. Berkeley [1685-1753] who, incidentally, can be refuted in his own terms by being simply thought out of having ever existed). The point here is that scientist A, who doesn't trust B's data, is somebody else's B and, thus, nobody's data are usable!

Finally, one can usually identify spurious data sets whether they consist of cleverly fabricated data or are due to some conceptual, recording or typographical errors (oceanographers and statisticians, indeed, write papers and books about such topics). This can be illustrated by Fig. 2, which includes a point — based on age estimates and gill data — for the coelacanth (*Latimeria chalumnae*), that is so far out that both growth and gill data appear to be erroneous. The gill data turned out to have been expressed as mm^2 instead of cm^2 (G.M. Hughes, pers. comm.). The growth data are also probably very wrong: it is hard to accept that *Latimeria chalumnae* should grow as fast as tuna!

Convinced?

If yes, there is only one more point to make: being ready to use other people's data implies publishing one's data in a form that others can use. This means that one should strive to publish as much as possible of one's raw data. Eloquent and well reasoned arguments for this have been made by numerous biologists, notably J.L. Munro (see Fishbyte 1986, 4(1):3). All of these authors emphasize the enormous waste, duplication of effort and loss of unique opportunities that nonpublication of costly experimental or field data entail.

The same waste, duplication and losses arise when one doesn't find ways of using a part of the vast trove of knowledge already in the literature to expand on or validate one's work. Thus, go ahead, use someone else's data!

On Using Other People's Data

Why Use Other People's Data?

DANIEL PAULY
ICLARM

Having received several encouraging letters following a previous essay on the fact that "Fisheries scientists must write" (see NAGA April 1986, p. 8-9), I shall try in the following paragraphs to deal with another aspect of a (fisheries) scientist's work that is rarely addressed explicitly: when and how to use other people's data.

What are the issues here? In order of importance, I would rank them as follows:

1. Further progress in fields such as fisheries biology, particularly in the tropics, will only occur if ways are found to turn the accumulated knowledge ("the literature") into an active body of information, rather than a repository of old bones.
2. Use of data generated by others is a compliment to, not a theft from, the persons who generated the data.
3. For (2) to be true, explicit rules should be made for students and young scientists on how data produced by others can legitimately be used.

Point (1) is rather straightforward: fisheries biology is (a) historical and (b) messy, and hence it is necessary to use lots of data -- often more than a scientist can collect during a research project of limited duration -- for interesting patterns to emerge. For example, assembling time-series data to test whether sunspots (indirectly) affect fish recruitment (Fig. 1) necessarily implies reliance upon other

people's data, since the working life of a research scientist, say, 30 years, could not cover more than 3 sunspot cycles: too few for any serious time-series analysis.

Fig. 2 shows how biological data are messy: it illustrates the thesis, developed further in some of my own technical papers, that the growth rate of fish is, other things being equal, a function of their respiratory (i.e., gill) area. Now this is a reasonable proposition which would be easy to test if biological data were not messy. However, they are, both because people often make awful mistakes, such as when estimating the gill area, and because "things" are not equal. Fish, besides breathing and growing, also feed and reproduce in various ways, resulting in a large scatter of data. The result of this in our example (Fig. 2) is that a stubborn scientist using only his or her data, painstakingly obtained from fish of the family Scombridae, would find no significant relationship between gill size and growth performance, while the more imaginative scientist, working on several fish families, would do so.

I stated above (in point 2) that using people's data is a compliment to their work, knowing very well that many feel very differently about this and, in fact, often become annoyed when it turns out that their data contained more information than they were able to extract in their (first) analysis. To these individuals, one can only respond that how much information is embedded in a given data set is a

function of history; claiming ownership to all (future) information that can be extracted from a data set is like claiming ownership of one's children: once released, they have a life of their own, and all one can claim, in a sense, is to have set them free.

Moreover, one must remember that the overwhelming majority of scientific papers never get cited (ever!), so the data they contain help nobody -- not even to make their authors better known.

Thus, for many of us, becoming a footnote in a much thumbed book, or a table entry in somebody's citation classic is likely to be the essence of our claim to fame and posterity.

Some Rules

For any of the preceding to be true, and for the use of someone else's data to be legitimate, a number of *rules* must be respected. I have perceived these rules through professional interactions with many colleagues, but never have I seen them written before (or maybe I have, but forgotten where). So share with me the effect they have when they are first stated:

Published material

- Use any numerical data as long as you cite the source.
- For parts of a text, put the material in quotation marks and cite source.
- If using a complete, published graph or table, cite the source and state that permission to reproduce the item has been obtained from the original publisher and/or author -- after you have obtained such permission.
- If using part of a published graph or table which is then incorporated into a new graph or table, or turning a table into a graph or vice versa, only a citation is needed.

Technically unpublished material

- Treat these the same way as the first group of items, except that

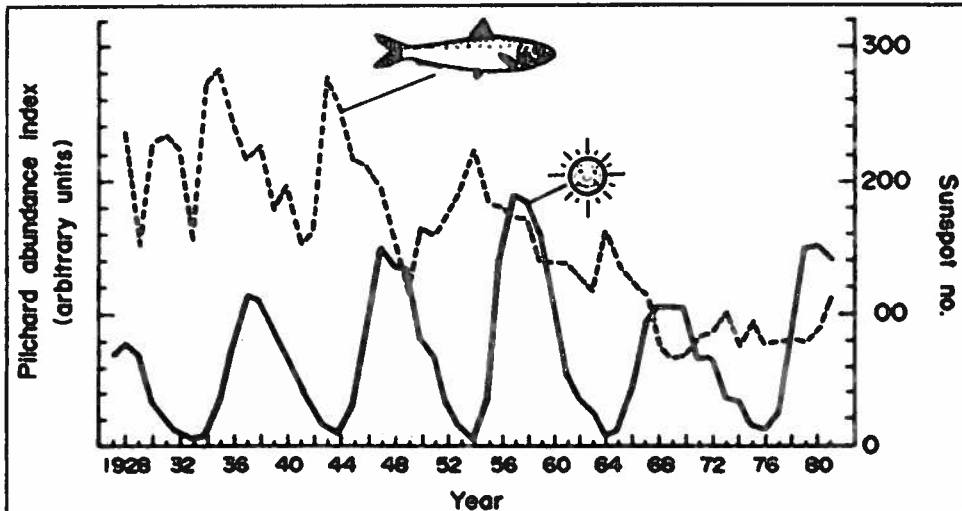


Fig. 1. Time series of sunspot numbers and of the Portuguese catch of *Sardina pilchardus*. Adapted from O. Moura and G. Afonso dos Santos (1984). I.C.E.S. C.M. 1984/H:48 Pelagic Fish Ctee, 11 p.