On Using Other People's Data

DANIEL PAULY
ICLARM

Why Use Other People's Data?

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.

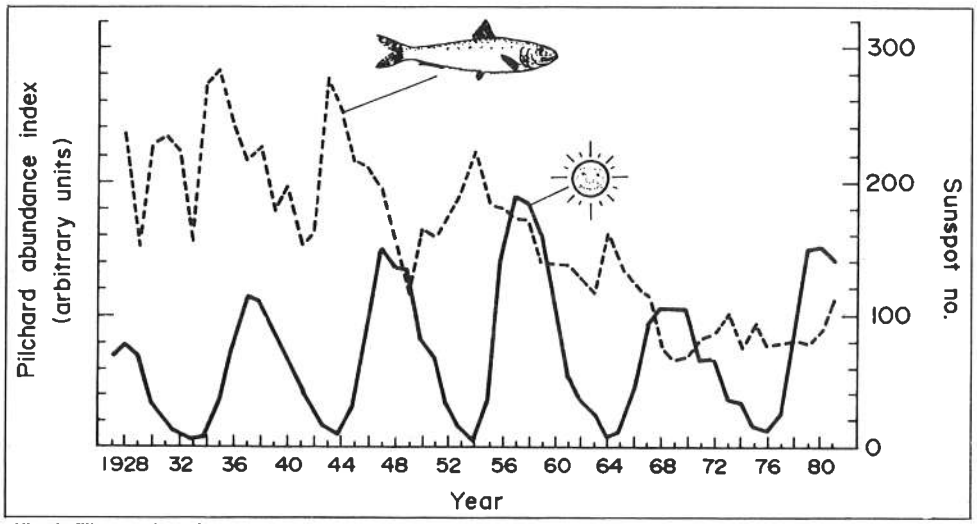

Fig. 1. Time series of sunspot numbers and of the Portuguese catch of *Sardina pilchardus*. Adapted from O. Moura and G. Afonso dos Santos (1984). I.C.E.S. C.M. 1984/H:48 Pelagic Fish Cttee, 11 p.

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