

How Good Are Your Data?

BY MATTHIAS TOMCZAK

ONE CHARACTERISTIC of universities that makes them unique as teaching institutions is the involvement of their teachers in research. The capability to draw on personal research experience allows lecturers to go beyond the presentation of results obtained by others. By showing examples of the scientific process at work, lecturers enhance students' ability to better judge the reliability and accuracy of results presented from scientific studies.

During the 1960s and 1970s, government initiatives to convince more people of the value of tertiary education led to a huge increase in student numbers. Many new universities were established with generous government funding. Today many governments no longer consider university education a part of the nation's culture. Instead, they treat universities as factories for the production of graduates. There is discussion of separating teaching from research by creating "teaching only" positions, or even "teaching only" universities. Research then becomes the "privilege" of a few selected universities, and the opportunity

Matthias Tomczak (matthias.tomczak@flinders.edu.au) is Professor of Oceanography, Flinders University, Adelaide, Australia.

to gain insight into the scientific process is no longer offered to the majority of students. Some countries have already gone down that path. But the generally accepted standard remains that true universities are based on a combination of teaching and research.

In oceanography, the pressure of large classes has not been as severe as in other disciplines, but other forces have produced similar results. Collecting data in the field provides ample opportunity for students to experience errors, difficulties, and instrument malfunction. Unlike data from controlled laboratory experiments, earth science data are collected under whatever conditions prevail at the time of the experiment.

Today fewer and fewer students experience what it means to collect data in a variable and not always friendly environment. Remote-sensing technology has opened opportunities to study the ocean without ever going to sea. Numerical modeling has become an integral part of oceanography, with an increasing number of oceanographers conducting all their research in an office on land. The need to understand and predict climate variability and change led to the compilation of data bases that cover the world ocean, extend over more than a century,

and are available for download on the World Wide Web. The web has made access to data easier than ever before. It allows lecturers to use ocean observations collected by others in any part of the world in their teaching and gives graduate students an opportunity to learn about the ocean without spending the time and effort to collect data themselves.

It is often said that practical experience is the best teacher. A student who never went out on a boat to collect data cannot fully appreciate the difficulties involved. Someone who worked through a night shift at sea knows that despite all good intentions not everything works smoothly at three o'clock in the morning: "Sorry, we'll have to repeat this station; I forgot to start the datalogger." "Can you please bring the instrument up on deck again? I think I forgot to remove the tube from the pump." "I was seasick and didn't realize that the cell was fouled up by a jellyfish."

Situations like these are familiar to anyone who has been on a research voyage. They shape your attitude toward data. You trust your fellow researchers and assume that they do the best job under the circumstances. But you also know that circumstances are not always favorable, that errors can occur, and that

not all data are of prime quality. When you find something unusual in your analysis of the data you ask yourself, “Could this be a problem with the data, or is it real?”

Observation shows that students who lack the experience of going to sea tend to trust all data blindly. In their eyes every data point is as good as its neighbor, and all data are of equal value. Pre-processed data such as climate databases are particularly seductive in this regard. By presenting data (sea surface temperature, salinity, oxygen, wind stress, heat flux, and much more) on a regular grid they give the impression of uniform data coverage and quality.

I remember an experience from the 1980s. One of my students had implemented an ocean general circulation model and was trying to model the global water masses correctly. He used sea surface temperature and salinity from the Levitus data set (the first version of today’s World Ocean Atlas) to constrain his model and could not produce Antarctic Bottom Water. I pointed out that although the Levitus data set is an excellent description of the mean state of the ocean, it is based on observations. Close to Antarctica it is strongly biased towards summer, as observations are rarely made under the winter ice. I suggested that he increase the winter salinity for just one line of data points along the Antarctic coast to simulate brine rejection during freezing. He implemented the idea in his model, et voilà, there was the Antarctic Bottom Water, spreading through his model ocean far across the equator.

In fairness to our students it has to be said that they are not the only ones who take uniform data quality in pre-

processed data for granted. Information on data statistics—data distribution in space and time, standard deviation for each grid element, data originator, etc.—is an integral part of every ocean data base, yet many papers that are based on pre-processed data make no mention that the authors checked how the data were originally intended to be used. Often this is of no consequence for the result, but that cannot detract from the

that operate on a drifting platform over a period of years is a challenge to ocean engineering. ARGO floats demonstrate that this challenge can be met, but individual floats sometimes return doubtful data. Making such data available in real time makes it more important than ever to develop in our students the ability to assess the quality of data collected by others. How can we do this?

When I went to university forty years

**Good science is impossible without good data.
Checking the quality of your data is therefore
as important as using the right tools for
their interpretation.**

fact that a brief statement on the suitability of the data set for the proposed research should be part of the paper. Ignoring such a basic premise provides a bad example for our students.

The issue of data suitability is becoming more prominent with increased accessibility of data in near real time through observing systems such as ARGO. ARGO floats, which provide measurements of temperature and salinity for the upper 1500–2000 m of the ocean, are extremely powerful observing platforms, and their virtually instantaneous data accessibility via the Internet marks a new era of ocean observation. It goes without saying that when using those data, careful assessment of data quality is required.

Providing uniformly high-data quality in near real time from electronic sensors

ago, oceanography had not received the public attention it enjoys today, and university classes in marine sciences were small. An integral part of the oceanography course was a short period at sea with the university’s research vessel, during which the students had to collect data along a transect across the western Baltic Sea. They then had to process the data and prepare a data report with a full description of the collection and quality-control procedures used. Some of the leading oceanographic teaching institutions of the world still maintain this standard of instruction. Faced with large class sizes and reduced government funding, most universities can no longer afford to expose their marine science students to this kind of experience at sea. (My own university is considering the sale of its small research launch, claiming

that it can no longer afford the \$5000 per year it costs to maintain!)

There is no substitute for hands-on experience. Shifting data collection from the sea into the laboratory is not the answer. To obtain data from an environment that the observer cannot control is the essence of earth sciences. The idea of the laboratory—to perform experiments under controlled conditions—has its rightful place in marine sciences as well; the properties of pure water and sea water, the effect of stratification, the conditions that lead to double diffusion, and many other processes are best studied in the laboratory. But collecting data at sea remains an integral part of every high-quality oceanography course.

To some degree, data collection at sea can be replaced by the collection of meteorological data on land. Seasick-

ness is not essential for the educational outcome, and in most other respects atmospheric data provide the same challenges to data collection and quality control as oceanic data. Given the growing importance of climate studies, which has led to closer collaboration between meteorologists and oceanographers, some exposure to meteorological data can only be beneficial to a marine science class. Wind measurements can be very instructive in that regard, particularly when the exercise is combined with the derivation of wind stress from the collected wind data: wind stress is a quadratic function of wind speed, and in gusty wind situations it makes a significant difference to the wind stress whether the wind is measured every second, every minute, or every hour.

The wide availability of pre-processed

databases requires additional focus. Individual data collected by generations of oceanographers remain the foundation of any regional or global database, and their value cannot be fully appreciated without personal experience of data collection at sea. But a complete assessment of the quality of pre-processed data requires an understanding of data statistics. An introduction into data quality control and some knowledge of regional data density should therefore be compulsory for any student who wants to embark on a career in oceanography. It should also be part of every undergraduate marine science course, where students are taught the essence of the scientific method. Good science is impossible without good data. Checking the quality of your data is therefore as important as using the right tools for their interpretation. ☒

Do you have a hands-on oceanography laboratory exercise to share?

We are interested in publishing peer-reviewed laboratory hands-on exercises along with our regular education column. We expect the text plus graphics to fit on two to four published pages. Text should include the lab's theme, a list of supplies needed, approximate time needed to complete the lab, steps students (and teachers) need to proceed through, and any useful references. Please clearly indicate the academic level of the lab (e.g., undergraduate, graduate). Submit text and graphics for review to the *Oceanography* Editor, Ellen Kappel at ekappel@geo-prose.com. More information on Hands-On Oceanography and published exercises can be found at <http://www.tos.org/hands-on>.