Appendix A: Finding data

Interesting data sets can be readily generated by integrating a differential equation. However, many students are much more interested in working on real data sets. It is surprisingly easy to obtain data for analysis.

Using an A/D board: By far the most exciting experience for science undergraduates is to analyze data that they collected themselves. Present day undergraduate teaching laboratories in biology, physiology and neuroscience have the ability to collect and digitize biological signals (e.g. Backyard Brain, Phoenix, AD Instruments). Using these devices it is relatively straightforward to obtain ASCII data files for heartbeats, breathing, electromyographic (EMG), and electroencephalographic (EEG) data for analysis. A popular laboratory experience was to record the signals generated by an electric fish (see Laboratory 9).

In addition a number of useful, open source Python packages have been developed by Andrew Straw (Institute of Molecular Pathology, Vienna) that enable researchers to use their personal computer to run experiments control an A/D board and hence collect data from multiple instruments. These software packages can be downloaded from

http://strawlab.org/

These include:

1. The PyUniversal Library which enables data to be collected through an A/D board

   https://code.astraw.com/projects/PyUniversalLibrary/

2. Vison Egg, a package for generating visual stimuli for psycho-physical experiments

   http://visionegg.org/
3. Motmot, a package for real-time collection and analysis of uncompressed digital images from a variety of courses

http://code.astraw.com/projects/motmot/

**Electronic circuits:** Many of the classic limit cycle oscillators of mathematical biology can be implemented as an electronic circuit. For example, Laboratory 5 shows how a circuit can be constructed to integrate the van der Pol oscillator. Data generated by this circuit can be collected using an A/D board and the results compared to those obtained using numerical integration.

**Internet data sets for analysis:** A variety of databases exist on the Internet that enable data sets to be freely downloadable. These include

1. PhysioNet provides access to large collections of recorded physiological data sets

   http://www.physionet.org/

2. Time Series Data Library, created by Rob Hyndman, provides about 800 time series from different sources

   http://robjhyndman.com/TSDL/

3. Google’s Trends provide data concerning google hits for search words as a function of time

   http://www.google.com/trends/

4. National Center for Biotechnology Information website provides sequences for all known genes and, in some cases, the complete genome of an organism


5. United States Census Bureau provides data sets for economic trends and demographics

   http://www.census.gov/

6. A variety of EEG data sets are available, including
7. Birdsongs in the form of *.wav files can be downloaded and analyzed in the frequency domain (see Laboratory 9).

http://www.ilovewavs.com/Effects/Birds/Birds.htm

8. Stock history files can be downloaded from Yahoo! finance using the Python program snip_stock.py.

Using a camera: Most students has access to a digital camera. The digital images are in the format of a *.jpg file. These files can be readily handled in Python as shown in Labs 4 and 9.

Using a scanner: A scanner provides another route to obtain data. In particular, a figure can be scanned, converted to a *.png image, and the digitized using the freely downloadable software package, Engauge

http://digitizer.sourceforge.net/