Chapter 4

Time

4.1 Case Study: River

What you need to know: the data are monthly average river flow level of Willamette River in Oregon. We use this data to demonstrate the effects of scale on perception of structure. Figure 4.1 displays the data plotted at two different aspect ratios. The right side plot shows flow against time in the typical aspect ratio for a time series plot, wide and short. Periodic trends are clearly visible corresponding to the yearly seasons. The last year (beginning at observation 120) appears to start with higher than usual flow. The left side plot shows flow against time in a 1:1 ratio. Global trends can be observed at this scale: there is a gradual increase in flow until time 60, and then a gradual decrease.

4.2 Case Study: UK Pig Production

When exploring multiple time series it is important to examine the lag relationships between series. This section discusses how the projection manipulation controls can be used for this purpose. The actions described in this example are actually familiar and well-established practices in exploratory time series analysis. The analogy to these electronic tools is to hold up the individual plots (that have been plotted on the same physical scale) to the light and slide the sheets of paper until the series peaks and troughs roughly match. Programming these physical occupations is not so difficult and has been available in specialist software for some time (Unwin & Wills 1988), but the point of this example is to demonstrate that the projection manipulation tools, surprisingly, can be applied for the same effect. This example demonstrates that the manual controls is a simple exploratory tool for multiple time series, that can be embedded in the framework of a fairly sophisticated environment for exploratory multivariate
Figure 4.1: Willamette river flow data at two different aspect ratios. Top plot shows 1:1 ratio (contracted in time), which reveals long term trends, such as the up then down. Bottom plot shows long time axis, revealing local seasonal trends.
Figure 4.2: The series indicator variable *Indic* is manipulated into the vertical projection, thus separating the series, and the indicator variables for the *Gilts* and *SB* series are manipulated into the horizontal projection, thus lagging the series on the other series.
data analysis. It is necessary to preprocess the data, but this can be easily automated for general multiple time series. (Indeed an S function is available on the web page for this article http://www.public.iastate.edu/~dicook/research/papers/manip.html.)

The data has 5 indicators of pig production in the United Kingdom, with measurements taken quarterly over a 12 year period from 1967 until 1978. The data can be found in Andrews & Herzberg (1985). The series are as follows:

**Series 1** Number of sows in pig for the first time, that is, a measure of intake into the breeding herd (GILTS).

**Series 2** Ratio of all-pig price to all-fattener feed price (PROFIT).

**Series 3** Ratio of sow and boar slaughter to total breeding herd size, that is, the removal of pigs from the breeding herd (SB).

**Series 4** Number of clean pigs (meat) slaughtered (CP).

**Series 5** Actual breeding herd size (HERDSZ).

The data is preprocessed in the following way:

1. Each series is standardized.
2. The 5 series are concatenated (Pigs):

\[ GILTS_1, \ldots, GILTS_{48}, PROFIT_1, \ldots, PROFIT_{48}, \ldots, HERDSZ_1, \ldots, HERDSZ_{48} \]

3. New variables are created:

   (a) Series indicator (Indic): \(1 \ldots 1 2 \ldots 2 3 \ldots 3 4 \ldots 4 5 \ldots 5\)

   (b) Lag indicators (5 variables)

<table>
<thead>
<tr>
<th></th>
<th>48</th>
<th>48</th>
<th>48</th>
<th>48</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gilts</td>
<td>(-1 \ldots -1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
</tr>
<tr>
<td>Profit</td>
<td>(1 \ldots 1)</td>
<td>(-1 \ldots -1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
</tr>
<tr>
<td>SB</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(-1 \ldots -1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
</tr>
<tr>
<td>CP</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(-1 \ldots -1)</td>
<td>(1 \ldots 1)</td>
</tr>
<tr>
<td>Herdsz</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(1 \ldots 1)</td>
<td>(-1 \ldots -1)</td>
</tr>
</tbody>
</table>

The **Indic** variable is used to separate the series by manipulating this variable into the vertical projection, leaving the series sequentially plotted vertically (Figure 4.2). Any one of the series can be lagged positively or negatively on the remaining series by manipulating the indicator variable for the series into the horizontal projection. For example, in Figure 4.2 the **Gilts** series has been lagged on the rest. We have stopped at a projection where the **Profit** peaks match the **Gilts** peaks. The lag relationship looks to be about 3 quarters, or 9 months. This is interpretable given that the gestational period is about 4 months and clean pigs are usually slaughtered between 4 to 6 months of age.
4.3 Case Study: TAO Buoys

4.3.1 Data description

The El Niño/Southern Oscillation (ENSO) cycle of 1982-1983, the strongest of the century, created many problems throughout the world. Parts of the world such as Peru and the United States experienced destructive flooding from increased rainfalls while the western Pacific areas experienced drought and devastating brush fires. The ENSO cycle was neither predicted nor detected until it was near its peak. The need for an ocean observing system to support studies of large scale ocean-atmosphere interactions on seasonal-to-interannual time scales was highlighted.

This observing system was developed by the international Tropical Ocean Global Atmosphere (TOGA) program. The Tropical Atmosphere Ocean (TAO) array consists of nearly 70 moored buoys spanning the equatorial Pacific, measuring oceanographic and surface meteorological variables critical for improved detection, understanding and prediction of seasonal-to-interannual climate variations originating in the tropics, most notably those related to the ENSO cycles. The first buoys were operational in March 1980, and more have been added since. This data comes from this array of buoys. Daily updates on the buoy recordings are at http://www.pmel.noaa.gov/toga-tao/home.html. The full data set that we formatted for use here has 178,080 measurements. This is large enough that working with this entire data set requires some of the tricks described in the chapter on large data. For this section, though, we focus on re-structuring the data in different ways, and looking at these subsets. The variables recorded in the data set are:

- Latitude, Longitude: Values returned by the buoy of its actual physical location
- Year, month, day: Time measurement taken
- Zonal Winds: Winds from the west, east
- Meridional Winds: Winds from the south, north
- Relative Humidity
- Air Temperature
- Sea Surface Temperature

4.3.2 Querying Drift

Figure 4.3 displays the location variables of a subset of 28 buoys in the southeastern Pacific. It is interesting to note that the buoys drift around some, some more than others. Using linked brushing we can query the points that indicate a buoy drifting from its mooring, and find out the day, month and year that this happened (Figure 4.3). It looks like someone was aware of the drifting and fixed the problems: all the time periods are consecutive and in the past.

4.3.3 Cycling over time

Variables were re-structured to have each variable correspond to a year’s monthly average measurements on air temperature for two different locations. Anima-
Figure 4.3: (top) Latitude vs longitude of buoys in the south-eastern Pacific, showing how some drift around a lot. (Middle, Bottom) Querying the drifting shows it to be consecutive days, and the drifting has been corrected in all cases.
tion over the year can be done to assess year-to-year trend. This is a common technique for handling time series. The plots in Figure 4.4 illustrate a static representation cycling through the monthly air temperature measurements for two different buoys over the years 1993-1998. The year 1993 was the year that the buoy at 180°W installed. (Missing values are plotted as the minimum value, 20°C. And the air temperature is plotted as raw values, not standardized.) Some interesting features are visible. The air temperature for the buoy further out in the ocean is generally higher than the buoy close to the coast. Typically air temperature at 110 starts off reasonably high and then dips lower. In contrast, air temperature at 180 tends to start lower and get higher. (The standardize scale would be better for assessing this relative trend.) The year 1997 is very interesting. The air temperature out to sea remains high all the year, and the air temperature near the coast continues to climb higher at the end of the year rather than decline. (This buoy stopped recording in May, June and July.) Indeed this trend continues on to 1998 also. This period marked an extreme El Niño event.

4.3.4 Correlation Tour

This section illustrates exploring multivariate time using the correlation tour. To do this we have taken the measurements made by one buoy 2°S and 110°W, including all wind, humidity and temperature variables. Monthly averages for each of the variables for two years (1993, 1997) were used. Cases with missing values on any variables were excluded from the plot, hence the break in the time series for 1997. The correlation tour allows for two separate 1D tours to be run on the horizontal and vertical axes using two different subsets of variables. In this example, we fix the horizontal axis to time, and tour on the 5 measured variables on the vertical axis. The correlation tour can be helpful in detecting multivariate temporal outliers, a point that is on the edge of two or more variables. In addition, the correlation tour may be useful in detecting relationships when two or more time series are displayed. Figure 4.5 displays three snapshots of the tour on the tao data. The top plot shows the average of the two wind variables against time, and the middle plot shows a contrast of the two wind variables plotted against time. The two years look fairly similar in both of these plots. The bottom plot shows an average of three variables sea surface temperature, air temperature and meridional winds. Here there is a very large difference between the two years in the latter half of the year, the start of an El Niño event. It seems that meridional winds play an important role in the event.

4.4 Case Study: Telephone Usage
Figure 4.4: Air temperature vs month for the years 1993-1998, for latitude 2°S, and (110°W,180°W).
Figure 4.5: Exploring multivariate relationships in time.