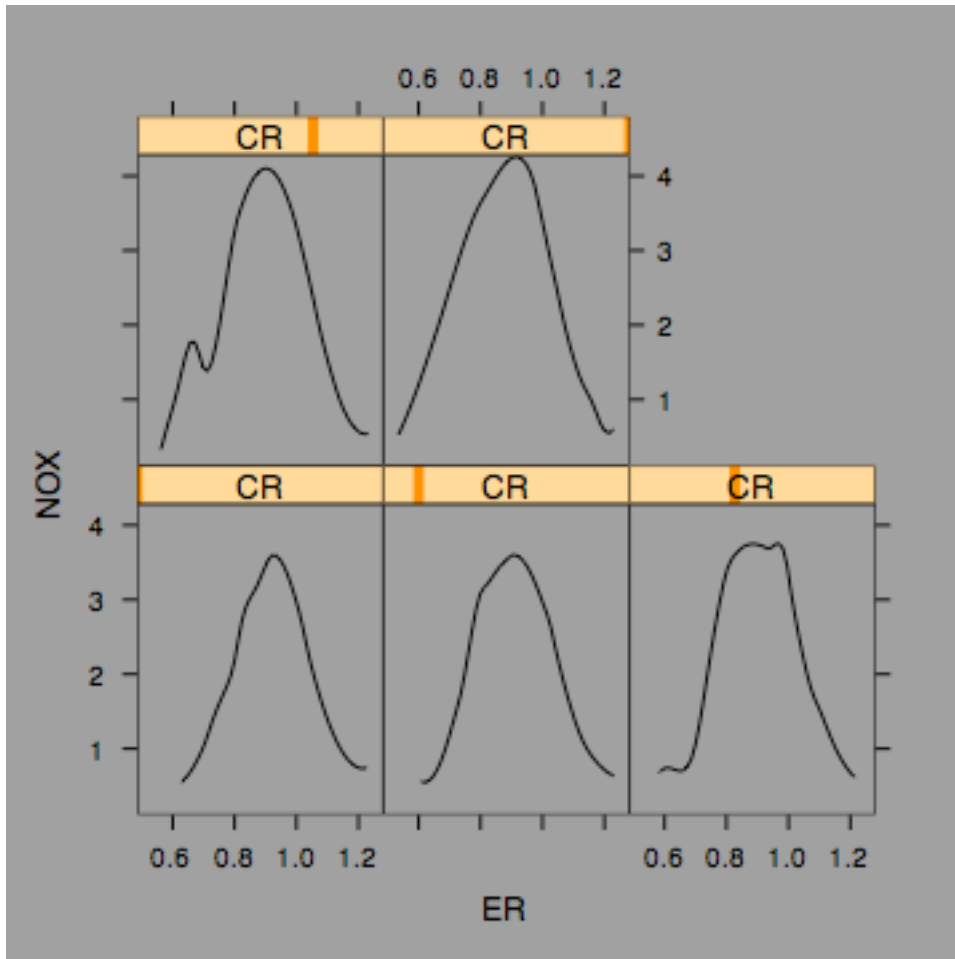


Ch 4: Level Plots of Data. The galaxy data.

But first ...

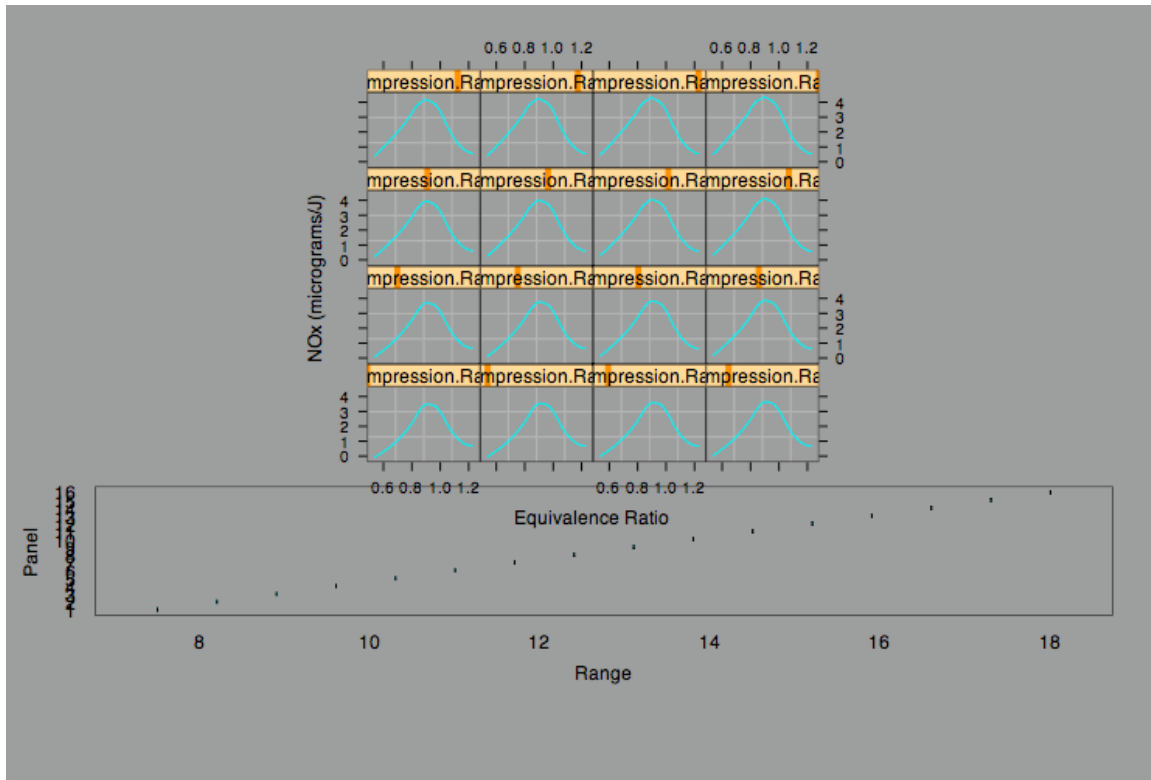
Controlling loess in coplots ....

```
xyplot(NOX~ER|CR,panel=function(x,y){panel.loess(x,y,span=.4,degree=2)})
```



Note that conditioned variable in this must be discrete. To do the other plot  $NOX \sim CR | ER$  we need the coplot command since it does the levels of ER automatically. Of course, if we discretize ER manually, we could use xyplot as above.

When a smooth surface is fitted to NOX based on a plane ER-CR, we get



This is produced from

```

> book.4.13 <-
+ function()
+ {
+   attach(ethanol)
+   eth.m <- loess(NOx ~ C * E, span = 1/3, degree = 2,
+     parametric = "C", drop.square = "C", family="s")
+   C.marginal <- seq(min(C),max(C),length=16)
+   E.marginal <- seq(min(E),max(E),length=50)
+   CE.marginal <- list(C=C.marginal,E=E.marginal)
+   CE.grid <- expand.grid(CE.marginal)
+   eth.fit <- predict(eth.m,CE.grid)
+   Compression.Ratio <- CE.grid$C
+   ans <- xyplot(eth.fit ~ CE.grid$E | Compression.Ratio,
+     panel = function(x, y) {
+       panel.grid(h=2)
+       panel.xyplot(x, y, type = "l")
+     },
+     aspect="xy",
+     layout = c(4, 4),
+     xlab = "Equivalence Ratio",
+     ylab = "NOx (micrograms/J)")
+   detach()

```

```
+ print(ans, position=c(0,.3,1,1), more=T)
+
+ ans <- plot(shingle(seq(min(ethanol$C), max(ethanol$C), length=16)),
+           sub = list("Figure 4.13",cex=.8),
+           scales=list(x=list(cex=.7),y=list(cex=.7)),
+           xlab = "Compression Ratio")
+ print(ans, position=c(0,.1,1,.45))
+ invisible()
+ }
```

More about this program resource shortly ...

But note the command

```
eth.fit <- predict(eth.m,CE.grid)
```

where eth.fit was a loess fit based on the original data.

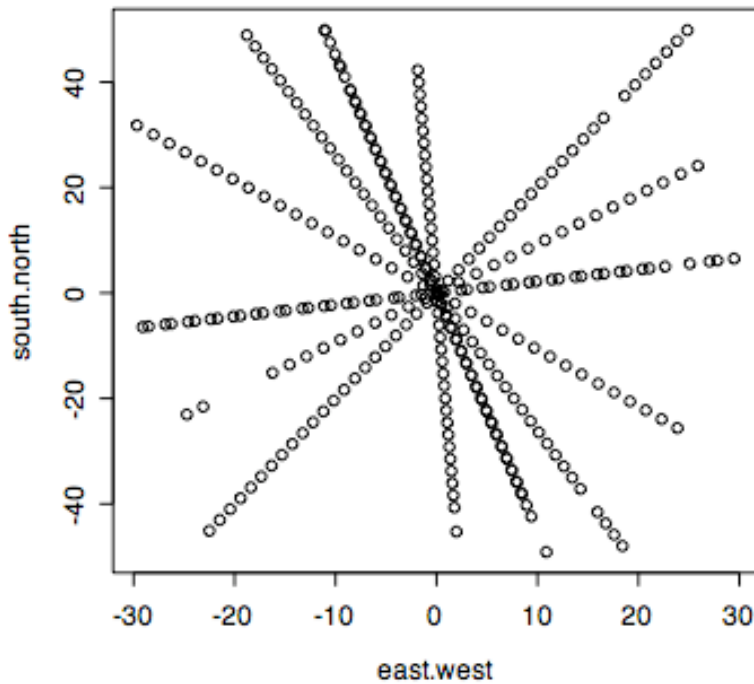
This is what generates the NOX values at the ER-CR grid (as opposed to observed NOX at observed ER-CR values)

**The galaxy data:** 415 measurements of the velocity of a galaxy. the direction of the sighting is carefully noted. (slit angle and radial position just give the same information in a different way.)

	east.west	south.north	slit.angle	radial.position	velocity
3	8.46278858	-38.17317200	102.5	39.099998	1769
4	7.96497774	-35.92769241	102.5	36.799999	1749
5	7.46716690	-33.68221283	102.5	34.500001	1749
6	6.96935558	-31.43673134	102.5	32.200000	1758
...	.....	.....	....	.....	.....
413	-9.09046364	41.00443268	102.5	-42.000000	1456
414	-9.58827496	43.24991226	102.5	-44.299999	1448
415	-10.04279804	45.30013657	102.5	-46.400002	1409
416	-10.54060936	47.54561615	102.5	-48.700001	1416
417	-11.06006432	49.88872528	102.5	-51.099999	1429

(Note the distinct feature of angular data – but not an issue here).

This is trivariate data but note the structure of the direction coordinates.



Imagine a spinning object, like a ball, that is receding from you. If you had a way of measuring its velocity along any line of sight, you could measure the velocity of each point on the ball at any instant. Would all such lines

of sight return the same velocity? No. At one instant parts of the rotating ball are moving towards you relative to the centre of the ball, and other parts are moving away from you. So the line-of-sight velocity measurement will be a bit more or less than the velocity of the centre of the ball.

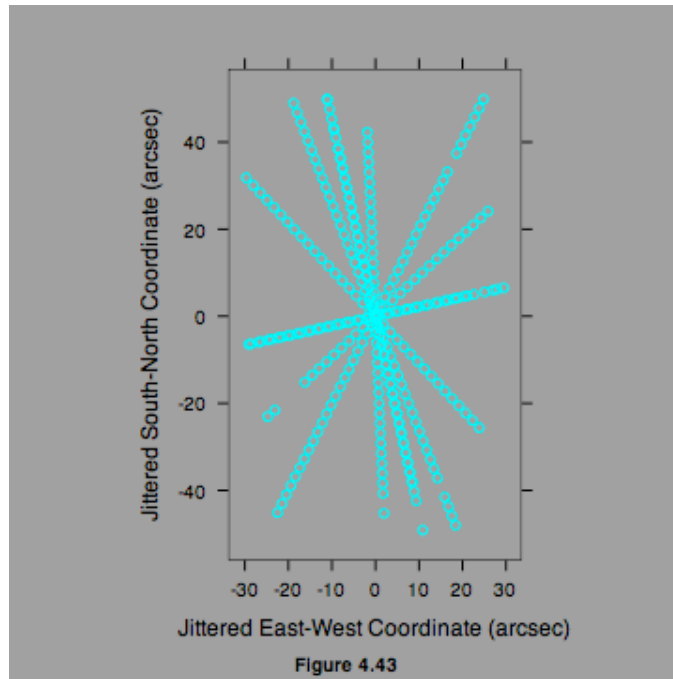
Now think of similar measurements of a galaxy – can we use the velocity measurements to detect the direction and speed of rotation?

Note east is to the left! Why?

Check out:

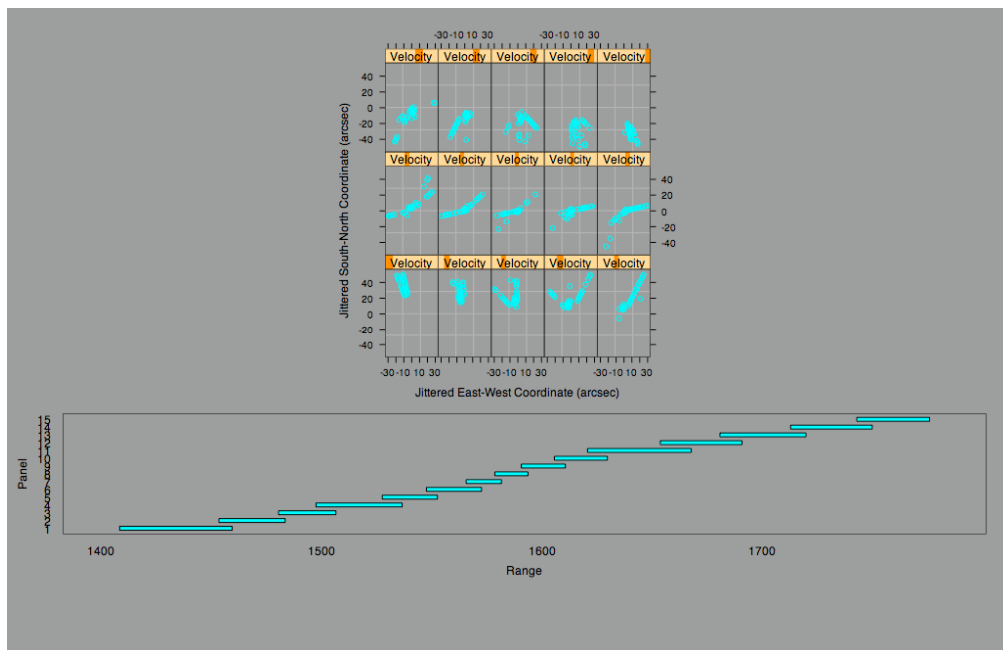
<http://cm.bell-labs.com/cm/ms/departments/sia/project/trellis/>

```
book.4.43 <-  
+ function()  
+ {  
+   attach(galaxy)  
+   set.seed(19)  
+   ans <- xyplot(jitter(north.south, .3) ~ jitter(east.west, .3),  
+               aspect = diff(range(north.south))/diff(range(east.west)),  
+               sub = list("Figure 4.43",cex=.8),  
+               xlab = "Jittered East-West Coordinate (arcsec)",  
+               ylab = "Jittered South-North Coordinate (arcsec)")  
+   detach()  
+   ans  
+ }  
book.4.43() produces .....
```

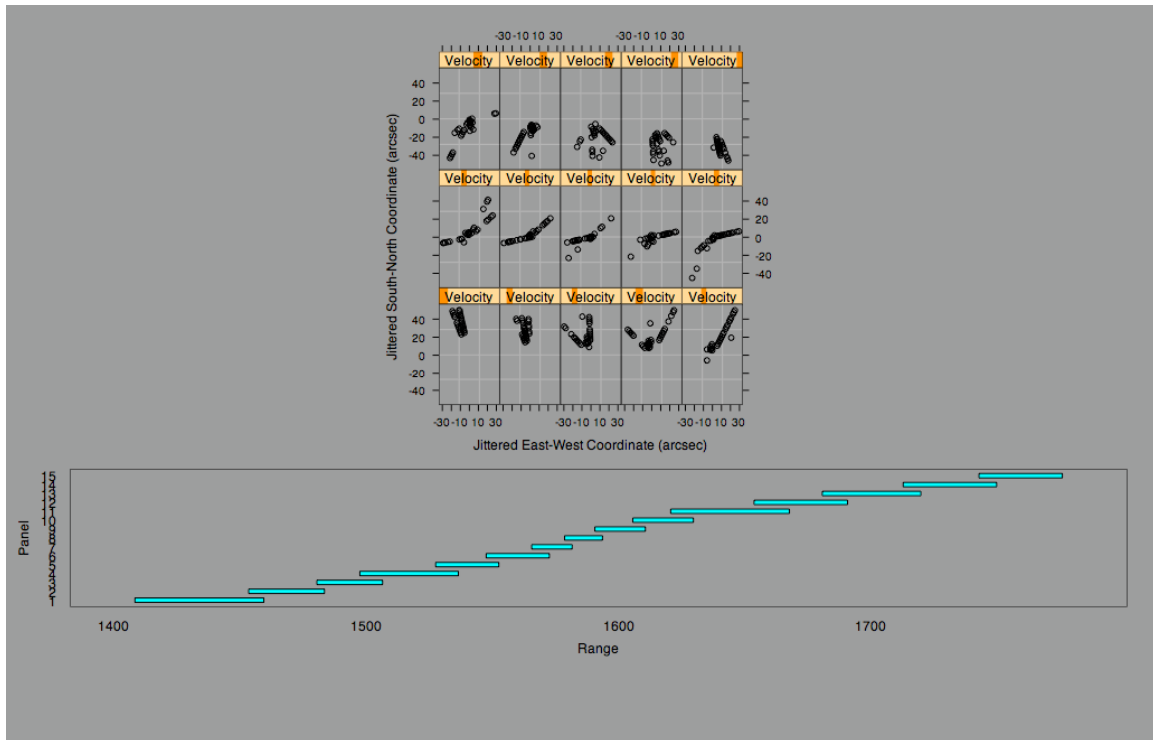


and

book.4.44() produces ....



Can modify programs ...



Very useful resource! (if you understand what Cleveland is doing ....)

At each point in this graph, there is a velocity measurement which might be represented as a coordinate measuring height above the page.

The velocities vary from about 1400 km/sec to about 1800 km/sec.

How can we graph this data? One way is with a “level plot”. See p 231.

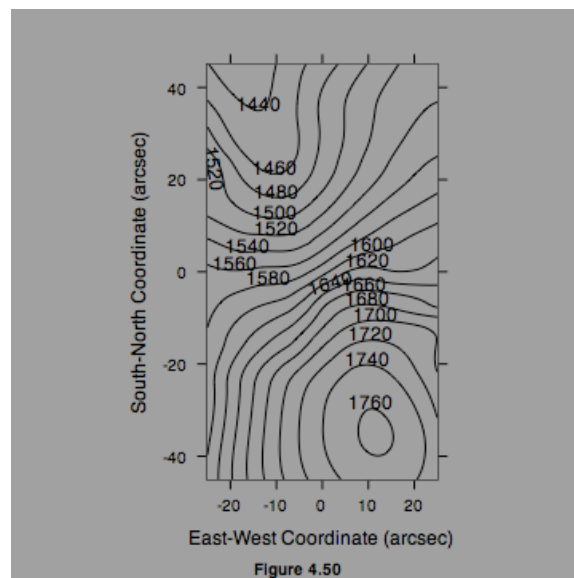
This is not a coplot – in which conditioning is usually on a factor or a measurement to be considered a factor. This is an instance where the E-W and S-N variables are not measurements but values selected by the experimenter, and the velocity is clearly the response, the dependent variable. Nevertheless, we can use the idea of conditioning to produce the level plot, by conditioning on the response.

The bottom left graph shows which directions lead to the lower velocities small on E-W and big on S-N, whereas the upper right panel shows where the high velocities come from (the opposite side of the galaxy). We can already tell which way the galaxy is spinning ....

The data is collected along concentric lines of directions (“slits” as Cleveland describes them). It is possible to describe each measurement in terms of polar coordinates – angle and radius. See p 233 for a coplot using of velocity against radius conditioned on slit angle. Can you see why these plots are consistent with the level plots?

What is Fig 4.46 on p 235. It is the same coplot as p 233 but based on the loess fit to the surface rather than individual loess fits to each slit’s data. However, the resulting residual plot is not further analysed.

Another way to look at this data is the Contour Plot.



This is slightly modified from the "book.4.50()" function. I had to delete the reference to "labels" and also the plot of the intermediate contours.

It is instructive to think about the mechanics of the contour plot. The data is not a smooth surface, but rather points on a rough surface. You need to smooth the surface first, and this is at a grid of points, not the data points. Then for each grid-square, you need to estimate where a contour line crosses this square. See pp 240-244 for a discussion of this. (This is important because you would not otherwise understand how the contour plot is constructed.)

Next : More Ch 4 pp 245 ff (A question to think about: What would a level plot of a smoothed surface look like?)