Sitka Spruce tree example

Exploratory Data Analysis

Let's begin by displaying the data. One thing we might do is simply plot the response trajectories for the trees for each chamber, each growing season. Figure 1 shows the logsize values for each chamber and growing season. The observed average at each day is also plotted and linearly interpolated *Figure 1: Observed logsize values (points) and averages (line) for each chamber and each year.*

To study the correlation structure, we need first to remove the effects of any explanatory variables, say the day and the experimental group (ozone treated or not)-we know from DLZ's analysis that chamber is not an important explanatory variable. I obtained the residuals from a 2-way anova model (OLS) on day and treatment group (with interaction) for the 1988 data.

Scatterplot matrix of residual pairs: To interpret this plot, the days should be evenly spaced. Here they are nearly so, with the difference between observation days ranging from 22 to 31 days. The correlations corresponding of these scatterplots are:

	Day152	Day174	Day201	Day227	Day258
Day152	1.00000	0.196862	0.262	−0.0321	0.002441
Day174	0.19686	1.000000	0.194	0.1993	−0.000278
Day201	0.26171	0.194033	1.000	0.2243	0.251451
Day227	−0.03209	0.199325	0.224	1.0000	0.264455
Day258	0.00244	−0.000278	0.251	0.2645	1.000000

The correlations between logsize values recorded 3 and 4 observation days apart appear quite small. However, the correlations at 1 and 2 observation days apart appear to be around 0.2. To estimate the autocorrelation function, we pool residual pairs that are the same number of observation days apart. For these 1988 data we obtain:

lag	1	2	3	4
$\hat{\rho(u)}$	0.220	0.238	−0.016	0.002
$se\hat{\rho(u)}$	0.056	0.065	0.080	0.113

The standard errors are computed as $1/\sqrt{N}$, where N is the number of pairs used when computing $\hat{\rho}(u)$.

Sample variogram for the 1988 data: We may also consider the variogram. Figure 3 shows the sample variogram values (u_{ijk}, v_{ijk}) for the 1988 data, together with the estimated variogram (the average of the vs at each u).

Maximum Likelihood estimation (DLZ 4.4)

Splus has a built in minimizer (nlmin) that we can use the maximize the likelihood function, once we write some code for the likelihood function. We discussed this in class and proper code must account for the restriction that V_0 is symmetric, positive definite, e.g. optimizing the Cholesky factor is an improvement over optimizing V0 using an unconstrained algorithm.

For the 1988 Sitka spruce tree data, eventually the Splus code for maximum likelihood estimation with unconstrained V_0 will converge, with reduced log-likelihood value -519.5 and

For a good, quick solution, use SAS PROC MIXED. See the SAS commands and SAS output for the Sitka spruce tree data.

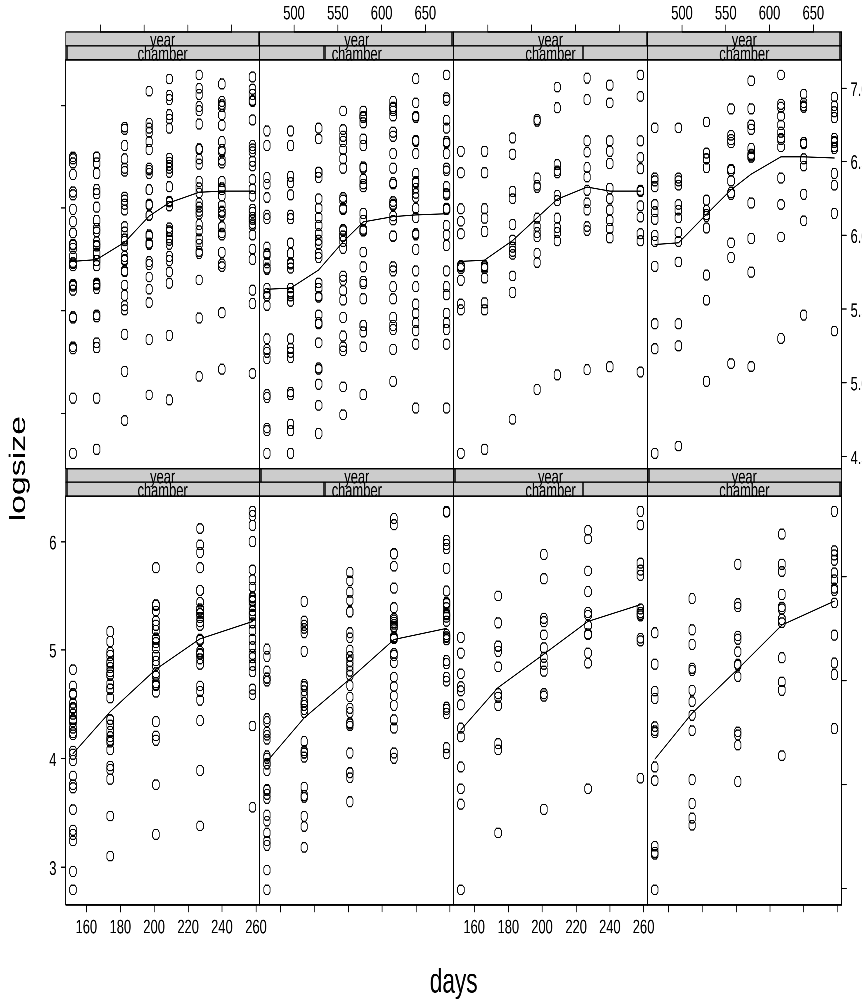

Figure 1: *Observed logsize values (points) and averages (line) for each chamber and each year.*

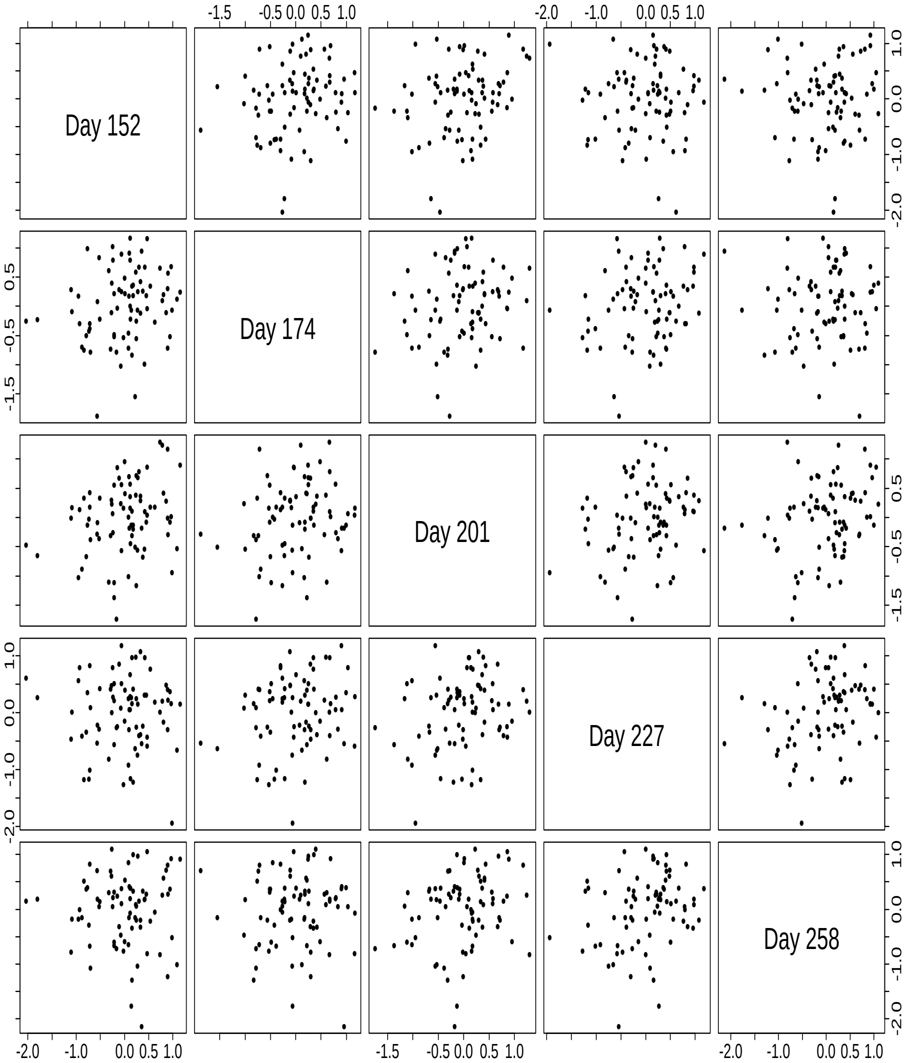

Figure 2: *Scatterplot matrix for the Sitka data set.*

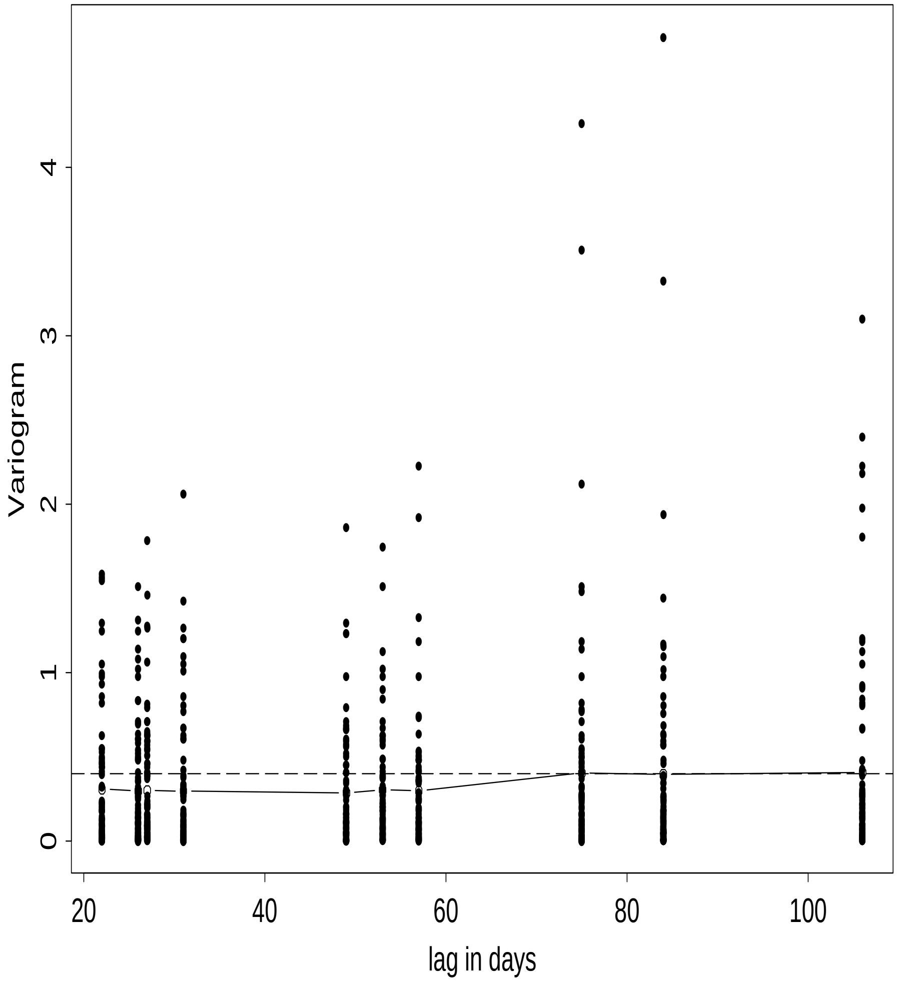

Figure 3: *Variogram for the Sitka data set.*