Homework 14

This assignment uses time series regression with generalized least squares. The dataset is a time series from a salt marsh in the Netherlands (Wadden Sea). The size/spread of a salt marsh is a complex balance of sedimentation, erosion, flooding, and plant growth and mortality. Mean high tide (MHT) fluctuates slightly from year to year, but because the marsh has very low relief this can have important consequences for physical and biological processes. If mean water level is lowered, or flooding is reduced (both are consequences of lower MHT), then the plant community may shift downward; likewise if MHT increases, the plant community may shift upwards.

This dataset records how three marsh zones fluctuate from year to year. The zones are Pre-Pioneer zone (sparse plant cover, but more than a mudflat), Pioneer zone (*Salicorna, Spartina*), and Marsh proper (*Puccinellia, Atriplex*). The column Distance is the location of the lower limit of the zone, in meters, relative to a seawall. A more positive number means the zone is located more seaward (also the numbers have been standardized by where the zone was in 1980). The column 'Zone' indicates which zone the measurement is from (PP = prepioneer, P = pioneer, M = marsh). 'MHT' is mean high tide, in millimeters, for that year.

The question to ask is whether MHT actually causes the location of the marsh zones to shift from year to year. Start by just analyzing the zone M. Make some exploratory plots of Distance over time, and Distance vs. MHT. What do the patterns look like?

Fit a linear model for Distance vs. MHT (just for zone M). Make an autocorrelation plot for the residuals of this model; does it look like the residuals are autocorrelated?

Fit a generalized least squares model instead, using an AR(1) model for the residuals. Does this model account for the autocorrelation? What is the estimate for phi? Does the GLS model affect your inference about whether MHT affects the height of the marsh zone?

Now we'll use the data from all three zones. If there is a general effect across zones, this should be better estimated by using all the data. But we need to account for autocorrelation *within* each zone's time series. Fit a GLS model to the data from all three zones, including predictors for MHT as well as the difference between zones in mean Distance. To model the autocorrelation within each zone, you can use AR(1) with a grouping factor, as described in ?corAR1. What does the model say about the effect of MHT, when pooling all the zones (you can use anova(model, type = "marginal") to get marginal F-tests for a gls() model)? How about the difference among zones? Does this model successfully account for any autocorrelation? To assess this you'll need to plot zone-specific acfs.

Maybe the effect of MHT differs among the zones; make an additional model that tests this.

Maybe we could account for temporal autocorrelation with a temporal smoother instead. Using just the zone M data, fit a GAM with a smoother for the temporal trend, while also including MHT as a predictor. What does the model look like? Does it successfully account for temporal autocorrelation?