mosqtemp meta data info

These data are time series of:

1) the average number of mosquito species trapped (SppRichness),

2) avg temperature (Temp)

3) urbanization (Urban)

4) total precipitation (Precip)

5) sediment DDT concentration (DDT)

**The data are collected for each year in a county in NY, from 1938 to 2012.**

From: Rochlin, I., Faraj, A., Ninivaggi, D.V., C.M. Barker, Kilpatrick, A.M . **2016**. Anthropogenic Impacts on Mosquito Populations in North America Over the Past Century **Nature Communications**7 13604 [Open Access Link](http://www.nature.com/articles/ncomms13604).

This paper asks: What are the drivers of changes in mosquito species richness?  The standard dogma is that climate change (i.e. temperature) is expected to increase mosquito abundance.  The set of mosquitoes in this dataset are mosquitoes with distributions to the south of the study area, so we'd expect the number of species to increase with increasing temperature.  But, because the data are time series, each data point may not be independent from the one (or more) before it.  The crucial question is whether the residuals of an analysis are independent (not the datapoints themselves - this is a frequent misunderstanding).

It's worth remembering that correlation in a response variable can come both from the variable being dependent on a past value, or because it depends on another variable X that shows correlation through time.  If you don't incorporate this variable X into the analysis the residuals will likely be autocorrelated and thus the assumptions of standard regression - that data points are independent (and identically distributed) is not valid.  You can deal with this by using an appropriate covariance structure that can account for the correlation created by dependence on unmeasured autocorrelated drivers or for the correlation of the variable with itself. Alternatively, if there is no inherent correlation in the response itself and if you can measure and incorporate the predictors that are resulting in the correlation in the response, the residuals will no longer be correlated.

    R functions that are helpful to assess correlation are pacf() which measures and plots the correlation of a dataset with itself at different lags, and also plots the confidence limits expected if there is no autocorrelation in the time series.

We'll be using the gls() function in package nlme to fit models with correlation.  I'd recommend reading the help file associated with this function.  You'll need to understand autoregressive models to understand how to specify the appropriate correlation structure for the gls model.

    It's worth noting that to extract residuals from a gls() object that take into account the correlation structure you've fit the model with you need to specify the type = "normalized" as in (e.g. for a fitted gls model called f1) :  
pacf(resid(f1, type ="normalized"))