Study site: Cheney Reservoir

* Large, shallow
* Eutrophic(average TP=100 µg/L)
* Cynobacteria-caused taste and odor and toxin events since 1990

Data set:

* May 2011 to June 2015
* Collected at the surface(0.5m) with Kemmerer sampler from May 2001 to July 2004. Vertical integrated photic zone samples were collected from August 2004 to June 2015.
* Geosmin: GC-MS
* Microcystin: ELISA
* Phytoplankton: membrane filtered slies

Data cleaning:

* Originally more than 100 physiochemical water quality variables were measured.
* Avoid collinearity: correlation greater than abs 0.75 were removed
* Explanatory variables with >5% of the observations missing were excluded
* Variables and response with concentrations less than the analytical limit of detection were substituted with a value half of the limit of detection
* Seasonality is an explanatory variable –Fourier transformed the data variable (i.e. sin and cos)
* 24 potential variables left
* Used elevation as surrogate for extreme precipitation events

Data cleaned:

* Dataset used to develop cyanobacterial abundance models: cyanobacteria\_abundance.csv(超链接):185 observations
* Dataset used to develop microcystin models: microcystin.csv(超链接):176 observations
* Dataset used to develop geosmin models: geosmin.csv（超链接）:185 observations