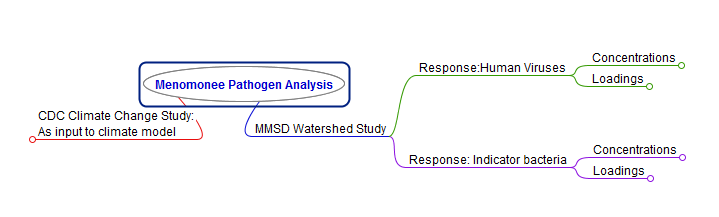
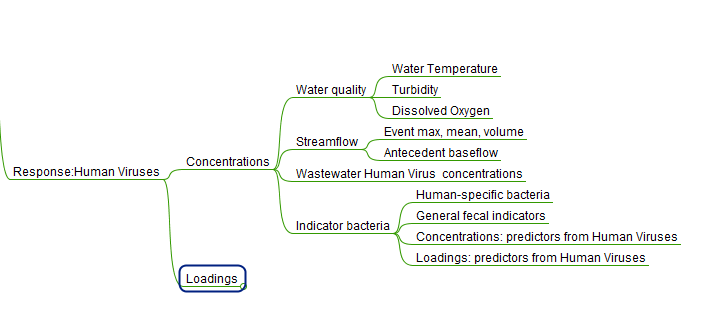
**First stage of analysis for MMSD virus results**

There will be several stages of analysis using data from the Menomonee River study that are illustrated in the chart below:



I. We will first be looking into the Human-specific virus concentrations (upper right node in the chart) as related to potential predictor variables. That node is expanded below:



Things to consider:

1. There are six stream sites. Most likely, the sites will behave differently. It is worth exploring a universal model for all six sites, but it will not be a surprise if relations with the predictors are better for individual sites.

2. Wastewater is the source, so the wastewater variables should play a part here. Viruses should be more likely to be present when they are abundant in the waste stream. It would be worth some discussion about how to include these variables most effectively. Possibilities: normalize response variable by wastewater equivalent (use ration of virus in stream to virus in wastewater as response??). Use wastewater as never-ever variable (if they are not in the wastewater they should not be in the stream).

3. It may be best to start with a model of human-specific FIB like BacHumancen rather than with viruses. There are less non-detects to deal with for the FIB data.

4. A thoughtful approach to variable selection will be needed with the number of predictor variables included.

5. A deliberate approach to the use of censored values in the response variables.

Notes from conversation with Mark Borchardt:

Review methods in 2011 paper on distribution systems in J of Water and Health

<http://www.iwaponline.com/jwh/009/jwh0090799.htm>

**doi:10.2166**.

Here is generally what they used for stats:

Response:

-concentrations (used gamma regression in bivariate regression)

-presence/absence (used binomial regression in bivariate regression)