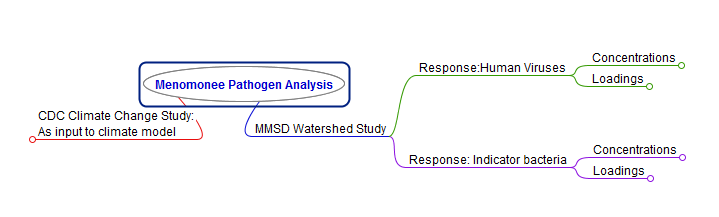
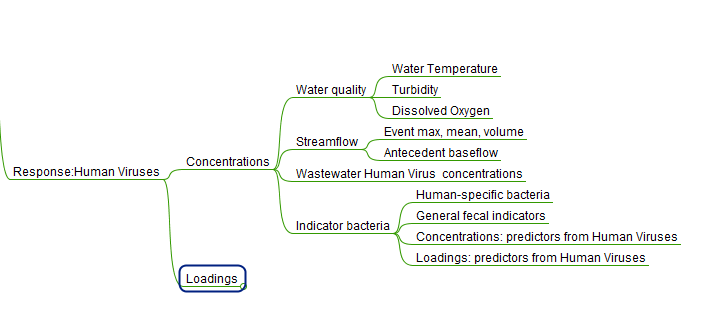
**First stage of analysis for MMSD virus results**

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:



**Objectives for first stage of analysis:**

* Establish relation between human-specific viruses and related variables including water quality variables and seasonality.
  + Turbidity, water temp, DO, streamflow, FIB, time of year, virus presence/magnitude in wastewater influent (human population)
* Establish relation between human-specific FIB (bachumancen, lachnocen) and water quality variables
  + Turbidity, water temp, DO, streamflow, time of year, (culture FIB??—ask Sandra), FIB magnitude in wastewater

The most useful outcome of this analysis would be to establish the probability of pathogen and human-specific FIB presence and relative magnitude under certain conditions.

**Data description:**

* 2 years of data
* 6 sites: Urban land use increases as the sites go downstream.
  + LM: Rural, tributary, Friestadt
  + MF: Rural + Low urban, main channel, Menomonee Falls
  + UW: Highly urban, tributary. Wauwatosa area
  + HW: Highly urban, tributary, Wauwatosa area
  + MW: medium urban, main channel, Wauwatosa area
  + MC: medium urban, main channel, downtown Milwaukee
* Baseflow samples every other month at a minimum. Every month for the second year for several sites.
* Runoff event samples 3 times per quarter throughout the study (rainfall, snowmelt).
* Underwood Creek (UW) and Memonomee River at Wauwatosa (MW) were not monitored during the entire study period due to interruptions in site access from channel or bridge construction.

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. For individual sites, the priority would be to analyze the downstream sites first since they have the highest level of occurrence. This is the order in which they should be analyzed using the site names and abbreviations within the data file:

MC, MW, HW, UW, LM, MF

2. Wastewater is the source, so the wastewater variables should play a part here. In theory, virus presence would be contingent on presence in the wastewater first. Presence in the wastewater will depend on presence in the human population. Magnitude in the wastewater depends on prevalence in the human population as well as dilution in the wastewater system. It would be worth some discussion about how to include these variables most effectively. Possibilities: normalize response variable by wastewater equivalent (use ratio 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. We might consider starting 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 considering the number of predictor variables included.

5. A deliberate approach to the use of censored values in the response variables is needed for the human virus data especially.

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)

Repeated measures analysis:

Variable selection: Choose which of the WQ variables that might be more useful for each category

Summarize why the variables are chosen—if more info on those, include it in the summary (why would something be seasonal)