MUDAC 2019
Problem Overview

Dr. Jonathan Hardwick, Computer Information Science Dept, MSU Mankato

This is real (and real messy) data

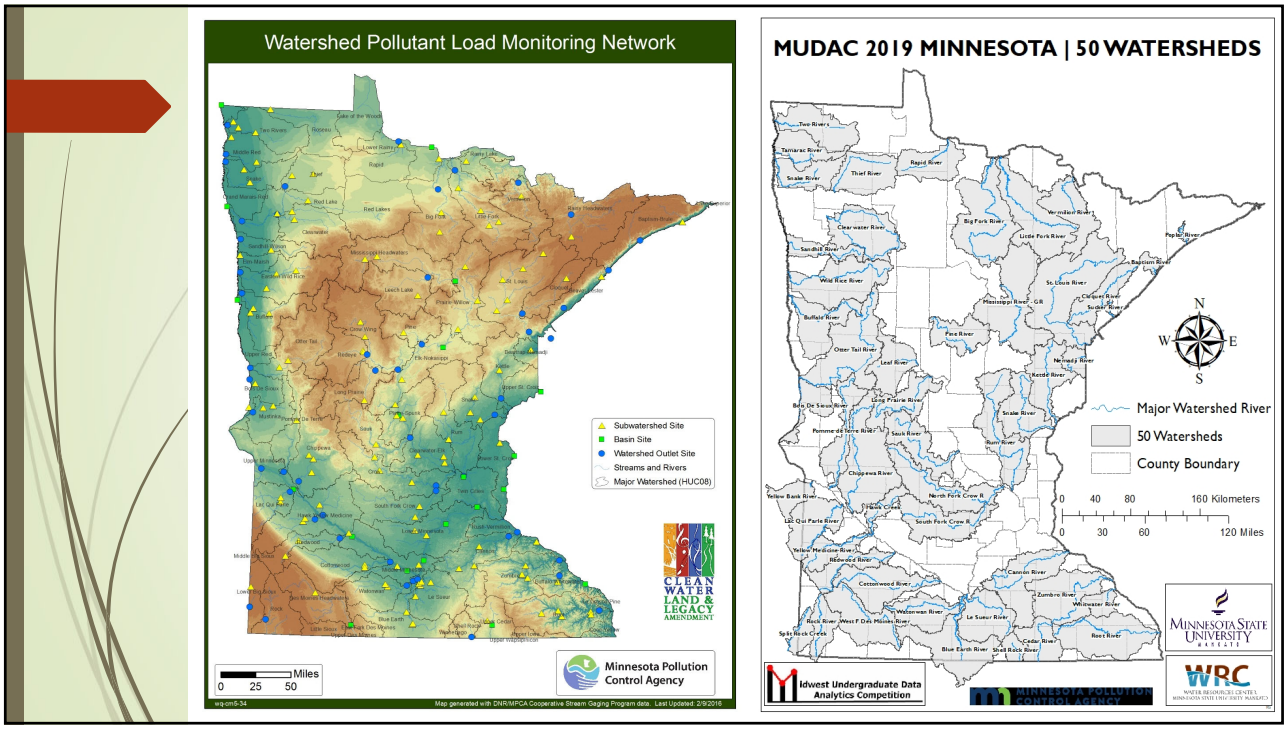
- The Minnesota Pollution Control Agency and Minnesota Dept of Natural Resources have funded a systematic water-quality monitoring program
- They have gathered a LOT of data
- Like much real data, it's messy, incomplete, and can be hard to interpret

These are real current open problems

- This is not a pre-canned homework problem that has a single solution
- Like many real-world problems, there is no "perfect" answer here
- Instead, we hope you can use your skills to provide insight and guidance
- For example, based on this data, what would you tell our policy-makers?

Your clients for this problem

- MPCA – Minnesota Pollution Control Agency
- WRC – Water Resources Center of Minnesota State University, Mankato

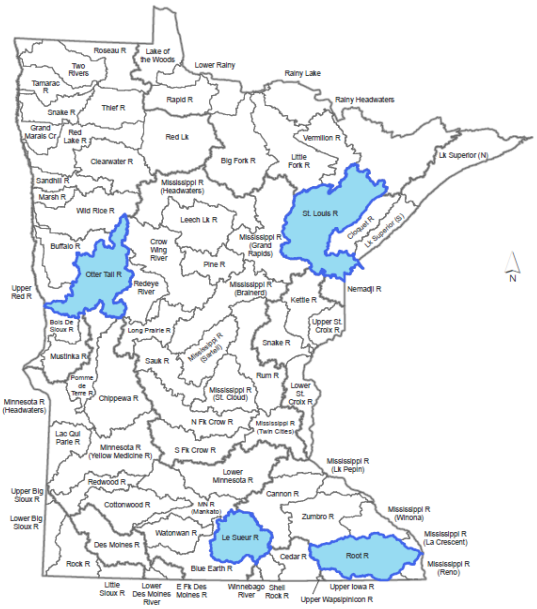


Problem 1: Build a Model

- We give you averaged data for 50 watersheds across Minnesota
- There are 15 watershed variables, and 2 water quality variables
- Build a model to predict the 2 based on the 15
  - Are the responses related to each other? If so, how?
  - Are the predictors related to each other? If so, how?
  - Do any watersheds seem to be outliers?
- A (relatively) constrained problem!

COLUMN	VARIABLE NAME	DESCRIPTION
WATERSHED VARIABLES		
B	FOREST[1]♦	% of watershed with forest cover
C	GRASS AND HAY[1]♦	% of watershed with grass or hay cover
D	WETLANDS[1]♦	% of watershed in wetlands
E	CROPLAND[1]♦	% of watershed that is cropped
F	DEVELOPED/URBAN[1]♦	% of watershed covered with houses, buildings, roads, parking lots, etc.
G	ALTERED STREAMS[2]	% of all streams in watershed that have been altered (e.g. straightened)
H	TILE-DRAINED LAND[3]	% of watershed that has artificial subsurface drainage. The term "tile" reflects that when this drainage practice first began, clay pipe was utilized.
I	SHALLOW BEDROCK UNDER CROPLAND[4]	% of watershed with bedrock (e.g. limestone, granite) that is not covered with much or any soil
J	SAND[5]★	Average sand % in watershed soils
K	SILT[5]★	Average silt % in watershed soils
L	CLAY[5]★	Average clay % in watershed soils
M	ORGANIC MATTER[5]	Average organic matter % in watershed soils
N	LAND SLOPE[6]	Average topographic relief (hilly-ness) of watershed
0	LAKES[1]♦	% of watershed covered by lakes
P	LAKE INTERCEPTION[7]	% of watershed land that drains to a lake that is at least 25 acres in size
WATER QUALITY VARIABLES		
Q	TOTAL SUSPENDED SOLIDS (TSS)[8]	The amount of suspended solids in the river expressed as a multi-year average concentration. Units are millgrams/liter. Sometimes referred to as suspended sediment.
R	NITRATE[8]	The amount of nitrate in the river expressed as a multi-year average concentration. Units are millgrams/liter. Nitrate is a dissolved form of nitrogen.

Watershed Location



Problem 2: Model Rainfall Intensity

- For problem 2 we've zoomed in to just four watersheds
- There is years of data for each, plus hourly rainfall data
- **Open research question**
  - What is the effect of rainfall INTENSITY (not just volume) on water quality?
- Build a model that lets you experiment with different ideas for "intensity"
- A very open-ended problem!

Questions?

- During the competition, come to the central office with your questions
- Common questions and answers will be posted on the competition blog
  - https://mudac2019.home.blog
  - It has an RSS feed, so you can subscribe to updates
  - Check the blog to see if we've already answered your question before asking it!
- We'll use the two data briefings (10pm and 7am) to summarize updates, keep you all moving forwards, and tell you about the judging on Sunday