
PROJECT PLAN: SURFACE WATER AND DRINKING WATER CONTAMINANTS

ABSTRACT

As part of a continuing effort to understand the status of usable water in the state of Kansas, the Kansas Data Science Consortium (KDSC), along with investigators from the ARISE Kansas project, is undertaking a project that aims to understand how contaminants in Kansan surface water sources affects contaminants in drinking water sources across the state. To do this, KDSC will leverage water quality data from both surface water tests and drinking water tests across the state, linking them so that we can understand how contaminants in surface water pose challenges for communities across the state.

1 INTRODUCTION

KDSC has maintained a water data dashboard for the state of Kansas for over a year. The dashboard gives a broad overview of the condition of water in the state of Kansas, focusing on four different topics: the High Plains aquifer, reservoirs, water quality, and extreme events (e.g., storms, floods, etc.). This project would be a continuation of the water quality dashboard, as we better understand how surface water affects drinking water quality. Currently the water quality dashboard just includes measurements of contaminants in water bodies across Kansas, but doesn't include a measure of the effects of these contaminants downstream in usable water.

This project aims to achieve two different goals. The first is to map communities to surface water sources that contribute to their drinking water. By including this on the water quality dashboard, we can have a deeper understanding of the importance of different surface water sources. For example, if a particular surface water source contributes water to the reservoirs of three different communities, a contaminant increase in this source is more concerning than in a water source that contributes to only one or even zero.

The second goal of this project is to understand how much contamination in surface water sources affects drinking water in the communities of Kansas. After initially mapping communities to surface water sources, we can then model how contaminant increases or decreases affects communities drinking water. This will then allow us to see where contamination problems are most common, which communities are least affected by contaminant changes in surface (and which communities are most affected), and potentially help to identify problem sources and communities that may require intervention.

2 BASELINE OR INITIAL ANALYSIS

Initial analysis should include:

- Connecting drinking water nodes with surface water nodes
- Testing that our connections are correct – can we double check using other sources?
 - o We want to know that some portion of the drinking water is coming from a particular surface water source. We can connect them via location and elevation changes, but we want to make sure we aren't missing anything and are connecting them correctly.
- Plotting out trends in contaminants in both drinking and surface water.
- Identifying most common contaminants.
- Identifying instances of contaminants exceeding EPA hazardous levels, looking at both drinking water and surface water tests.

3 FINAL ANALYSIS

The final analysis will involve taking the mapped communities and surface water sources and analyzing the data in several different ways:

- An overall simple OLS model that takes the contaminants level in a community as the dependent variable and the contaminants level in a surface water source that has been mapped to that community as the independent variable. This model would be a simple one:

$$Drinking_{mt} = \beta_0 + \beta_1 * Surface_{st} + Contaminant_{st} + \epsilon_{smt}$$

where $Drinking_{mt}$ is the level of contaminant in the drinking water for a municipality (m) at time (t), $Surface_{st}$ is the level of contaminant in a particular surface water source (s) at time (t), and $Contaminant_{st}$ is a series of dummy variables for each type of contaminant measure. This will get at the overall effect of a particular contaminant change in the average surface water source on the drinking water that it has been mapped to. This analysis will be on the contaminant, surface water, and year level.

- Using the model above, we can see which contaminants are most likely to pass through to the drinking water, using the contaminant dummy variables.
- A map of surface water sources and the communities that they connect to.
- A plot of contaminants in a surface water source and contaminants in the drinking water sources that they connect to over time.

4 FINAL GOALS & EVALUATION

- Analysis of how much surface water contamination matters for community drinking water.
 - o Are our procedures for cleaning surface water adequate enough? Where are they clearly not adequate?
 - o Which contaminants are most likely to be passed through to the drinking water? Which are least likely to?
- Analysis of surface water location and drinking water linking:
 - o Where are drinking water sources pulling from? Which are the most important surface water sources to monitor?

5 RELATED WORK

- <https://waterhub.ku.edu/kansas>.

6 DATA & TECHNICAL REQUIREMENTS

Provide expected technical requirements and relevant datasets so the course staff can best support your project needs. What software libraries do you plan to use? What datasets are you using or need access to? Anything else we (the course staff) should know?

Relevant datasets have already been uploaded to ResFS. Software needed includes Stata, R, and ArcGIS. We may potentially need to chat with someone who is more experienced with GIS in order to understand how best to connect surface water sources with drinking water sources. Elevation and terrain will be important factors here, which R and Stata will have trouble accounting for.