geoconvention Virtual Event September 21-23 2020
CSPG • CSEG • CWLS • GAC • MAC • IAH

Mining of a Large Dataset to Investigate *E. coli* Contamination in Private Drinking Water Wells

Katie J. White[1], C. Schuster-Wallace[2], S. Dickson-Anderson[1], P. Hynds[3], S. Brown[4], K. McDermott[5], and A. Majury[5,6]

[1]Department of Civil Engineering, McMaster University, [2]Department of Geography and Planning, University of Saskatchewan, [3]Irish Centre for Research in Applied Geosciences, University College Dublin, [4] Department of Chemistry, Queen's University,[5]Public Health Ontario, [6]Department of Biomedical and Molecular Sciences & School of Environmental Studies, Queen's University

Summary

While all groundwater wells are susceptible to contamination, private drinking water wells in Ontario are unregulated and therefore individual well users are responsible for stewardship of their own drinking water. Climate change coupled with land use changes have been altering these risks through changing contamination levels and timing, which increases risks for the unregulated user. Data mining identified relevant spatial and temporal relationships between *E.coli* concentration, well characteristics, seasonality, and geographic location within a large dataset. These findings are useful for the creation of an improved fate and transport tool that increases contamination awareness of private well users.

Theory

An estimated 33% of the Canadian population relies on groundwater for their domestic water supply, including 70% of rural residents (Rivera, 2017). Of these, 12% (4.1 million) rely on privately owned and maintained groundwater supplies (Murphy, et al., 2016) that are outside government regulation and oversight. Historically, approximately 66% of Canadian waterborne disease outbreaks originated from privately owned wells or small municipal systems (Schuster, et al., 2005), which may be due to limited resources for maintenance, management, and protection strategies (Rivera, 2017).This study demonstrates the utility of a large, integrated dataset (Reichstein, et al., 2019), and identifies critical factors impacting the contamination level of *E. coli* (a fecal contaminant indicator) in private drinking water wells (Figure 1).

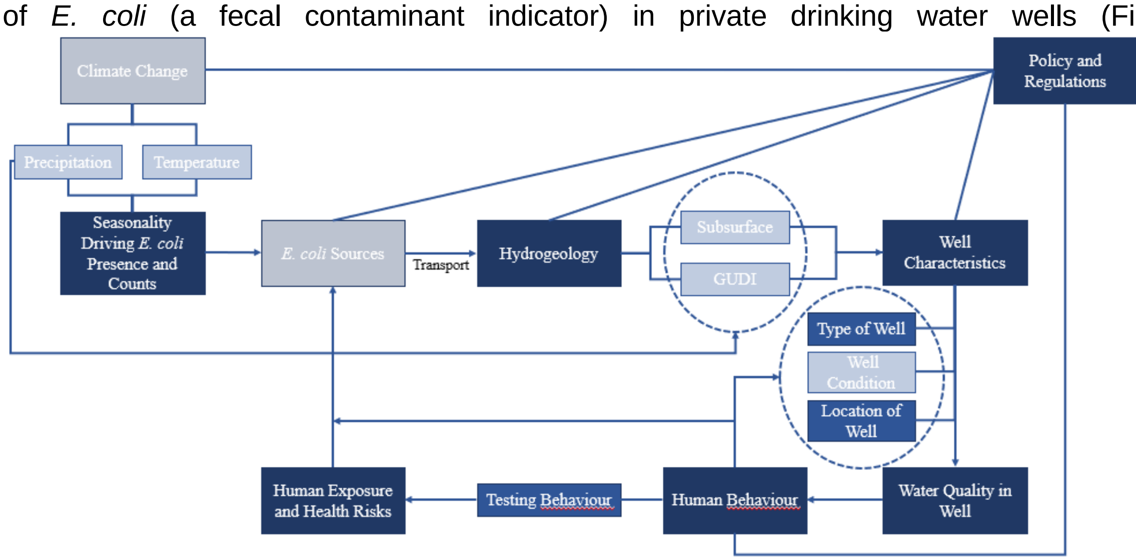

Figure 1: Exposure pathway of pathogens in existing private wells considering a coupled-systems approach, adapted from Di Pelino (2019). Blue boxes represent factors explored in this work; grey boxes represent future work. GUDI: groundwater under direct Influence of surface water.

Method

This work aims to contribute additional knowledge to select points along the fecal contamination exposure pathway depicted in Figure 1. The dataset consists of 795,023 observations for 253,136 unique private wells captured from 2010 to 2017. A variety of statistical and machine learning techniques were utilized to identify relationships between *E. coli* contamination and parameters such as well depth, time of year, stratigraphy, casing diameter, screen presence and details, and geographic location. Methods utilized include association rule analyses using the "apriori algorithm", univariate and bivariate analyses, regressions analyses, and variable discretization techniques.

Results

Specific relationships identified that *E. coli* contamination levels are greater in shallow wells, larger casing diameters, and consolidated formations. Within consolidated formations those in sedimentary and igneous type bedrock have a greater probability of higher concentrations of *E. coli* compared to those installed in metamorphic formations. The results of this data mining will inform in the creation of a user-accessible fate and transport tool by determining the most relevant relationships between *E. coli* and other previously mentioned parameters, that need to be investigated further and defined explicitly based on this dataset.

Acknowledgements

This work is supported by the Canadian Institutes for Health Research, and the Natural Sciences and Engineering Research Council of Canada.

References

Di Pelino, S., Schuster-Wallace, C., Hynds, P., Dickson-Anderson, S., & Majury, A. (2019). A coupled-systems framework for reducing health risks associated with private drinking water wells. *Canadian Water Resources Journal, 44*(3), 280-290.

Murphy, H., Thomas, M., Schmidt, P., Medeiros, D., McFayden, S., & Pintar, K. (2016). Estimating the burden of acute gastrointestinal illness due to Giardia, Cryptosporidium, Campylobacter, E. coli O157 and norovirus associated with private wells and small water systems in Canada. *Epidemiology and Infection, 144*(7), 1355-1370.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 195-204.

Rivera, A. (2017, January 10). *"The state of ground water in Canada* | *Ground Water Canada.* Retrieved November 2017, from Ground Water Canada: https://www.groundwatercanada.com/education/the-state-of-ground-water-what-we-know-and-dont-know-about-ground-water-in-canada-3584

Schuster, C. J., Ellis, A., Robertson, W., Charron, D., Aramini, J., Marshall, B., & Medeiros, D. (2005). Infectious disease outbreaks related to drinking water in Canada, 1974-2001. *Can J Public Health, 96*(4), 254-8.