Effects of Spatial Assumptions

Characterizing Tap-to-Household Water Contamination in India

Introduction

People in developing countries often have to collect water from public tabs or tanks. Even if the water quality at the source is adequate, there is a risk of contamination during transport to each individual household. Extensive water quality monitoring between sources and households is usually not feasible because of cost and privacy concerns. Thus, available water quality data tends to be limited in terms of temporal and spatial distribution. This project focuses on developing methods to characterize tap-to-household water contamination in locations with limited data availability. Comparison of multiple spatial and temporal assumptions of large data sets requires significant time and effort. This project used ArcGIS and statistical software (RStudio) to automate the majority of the analysis. The resulting tools and methods will allow for easy and fast replication of this analysis with similar data.

Testing of the methods and tools was performed using water quality (pH, nitrate, TDS, total coliform [TC], and fecal coliform [FC]) data from 160 households and 60 public water taps in two urban slums in Vellore, Tamil Nadu, India. This data set is considered limited for a tap-to-household contamination study because:

- sample collection did not follow any spatial or temporal pattern, and
- the water source for each household is unknown.

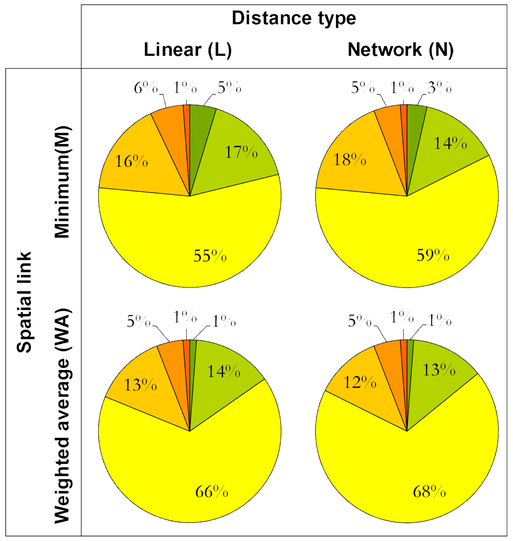
These two limitations translate to assumptions regarding the spatial and temporal connection between water quality at the tap and at the household.

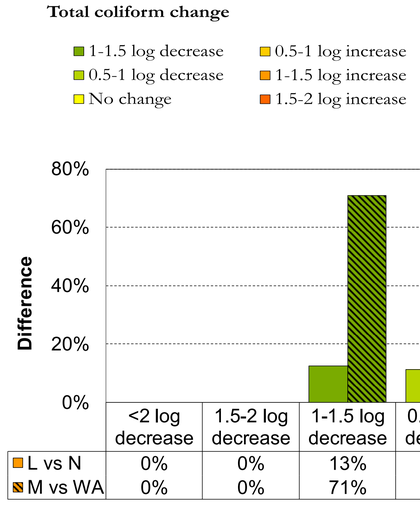
Tap-to-Household Contamination
Vellore, India

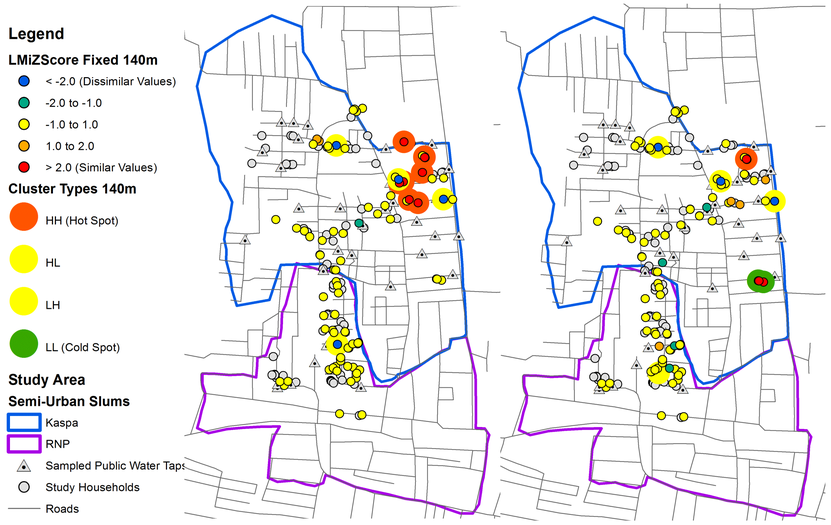
Comparison and Effects of Spatial Assumptions

Comparison of Spatial Assumptions

Changes in total coliform between tap and household were calculated based on distance type (linear & network) and on spatial link (minimum & weighted average). The pie charts represent the percentage of households in each category of Total Coliform (TC) change. The chart shows that the spatial link assumption has a similar or greater effect on TC change than the distance type assumption.





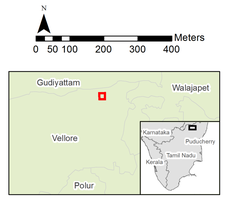


Spatial Link: Distance Weighted Average

Spatial Link: Minimum Distance

Effects of Spatial Assumptions

Hot/cold spot analysis was performed to exemplify the effects of spatial assumptions on data analysis. Only spatial link assumptions were tested since they showed the greatest effect on TC changes. The figures show a clear difference in the location and quantity of hot and cold spots based on the spatial link assumption.



Tania Alarcon
May 6, 2015
Projection: : Transverse Mercator (WGS 1984 UTM Zone 44N)
Data sources: Indo-US Study Data; GADM database of Global Administrative Areas; OpenStreetMap

Methods and Tools

1. Calculate linear distances using Near Table, join with tap and household water quality data, and save as a csv file.
2. Calculate network distances using Network Analyst and OpenStreetMap roads layer.
3. Edit the Network Table generated in step 2 to include household and tap IDs, and save as a csv file.
4. Load and merge output files from steps 1 and 3, select a subset of data based on a temporal assumption, and save data set.
5. Load output file from step 4, calculate tap-to-household changes for the water quality parameter of interest, select a subset of data based on a spatial assumption, and save as a csv file.
6. Perform hot/cold spot analysis for tap-to-household contamination using Cluster and Outlier Analysis (Anselin Local Moran's I) tool with a fixed distance band.

Step	Software	Automated	Tool		
			Name	User input	Limitations
1	ArcGIS	Yes	LinearDistance	Input file Output file	Long processing time
2	ArcGIS	No	--	--	--
3	ArcGIS	Yes	NetworkDistance	Input file Output file	--
4	RStudio	Partially	TemporalAssumption	Input file Output file	Only tests same-day assumption
5	RStudio	Partially	SpatialAssumption	Input file Output file WQ parameter	Only tests minimum & weighted-average assumptions
6	ArcGIS	Yes	HotColdSpotAnalysis	Input file Output file Distance band	--

Results and Conclusions

Spatial assumptions affect changes in tap-to-household contamination patterns for total coliform. There was no clear benefit from calculating network distance in these urban areas. Rural areas may show different results, thus future work will replicate this analysis on rural study sites.

The calculation method used for the tap-to-household link had a significant effect on total coliform concentration change. The distance weighted average method resulted in more locations with no change in total coliform concentrations than the minimum distance method. However, a higher number of hot spots were identified using the distance weighted average method.

Testing of assumptions is essential for characterizing tap-to-household water contamination. The methods and tools developed as part of this project allow testing of different spatial assumptions without having to repeat multiple time-consuming steps. Outputs from these tools can be used to generate multiple GIS layers that when combined will show hot spots of contamination regardless of temporal and spatial assumptions. These maps can then be used to study risk factors of contamination changes in the study area.