Diving into Bear Creek

ASA Colorado Wyoming Chapter Spring 2016 Meeting

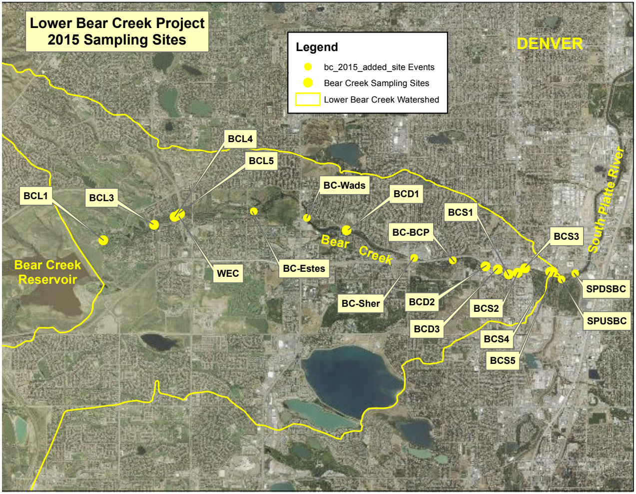
Presented by:
Caitlyn Cole, Ian Greenwald & Cody Griffith

Metropolitan State University
of Denver

April 15th, 2016

The Problem

Bear Creek runs from Morrison and meets with the south Platte River in Englewood. Recently, larger than normal populations of E. coli have made the river dangerous!



The Problem

This bacterium is related to the digestive track of humans and animals. E. coli is known to cause:

- severe food poisoning.
- flu-like conditions (fever, vomiting, etc.).
- in extreme cases, death.

Having an E. coli population count under 126 is considered safe by the EPA. Certain parts of Bear Creek have been measured to contain over 2420!

Prevention!

Reducing and preventing the E. coli populations in a local river will keep the public safe and increase access to usable water.

The Problem

This bacterium is related to the digestive track of humans and animals. E. coli is known to cause:

- severe food poisoning.
- flu-like conditions (fever, vomiting, etc.).
- in extreme cases, death.

Having an E. coli population count under 126 is considered safe by the EPA. Certain parts of Bear Creek have been measured to contain over 2420!

Prevention!

Reducing and preventing the E. coli populations in a local river will keep the public safe and increase access to usable water.

Along with the help of Groundwork Denver and our Biology department, we have begun to address the situation in three ways:

- We have formed a Multiple Regression model to incorporate a series of variables like: pH, turbidity, and location of sample.
- Created a Logistic Regression model to actually address the probability of danger given certain conditions.
- Incorporated spatial and temporal aspects into a kriging model.

Through different approaches, we should see some agreement and perhaps different insight into what is causing this problem.

Models

Along with the help of Groundwork Denver and our Biology department, we have begun to address the situation in three ways:

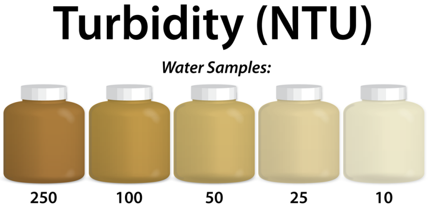
- We have formed a Multiple Regression model to incorporate a series of variables like: pH, turbidity, and location of sample.
- Created a Logistic Regression model to actually address the probability of danger given certain conditions.
- Incorporated spatial and temporal aspects into a kriging model.

Through different approaches, we should see some agreement and perhaps different insight into what is causing this problem.

Variables being considered

Variables:

- Turbidity (NTU) - The cloudiness of a fluid caused by particles being suspended within. Our values ranged from 1-40.



- Acidity (pH) - Our values ranged from 6.44 to 9.29.
- Sample Taken - At bank or in stream.

Metropolitan State University of Denver

Other variables

Things we found to be insignificant or didn't have enough measurements were:

- Dissolved Oxygen
- Flow Rate
- Temperature
- Shadiness
- Precipitation
- Others...

Regression stuff

We used multiple regression to make models that would predict E. coli levels.

West Model:

predictors = pH, log-turbidity

response = log-E. coli

Select statistics for this model:

df=156

$R^2 = 0.2067$

p-value= $5.327 \cdot 10^{-9}$

East Model:

predictors = pH, log-turbidity, interaction between pH and log-turbidity, where sample was taken

response = log-E. coli

Select statistics for this model:

df=387

$R^2 = 0.08883$

p-value= $4.162 \cdot 10^{-8}$

Predicted E. coli counts (west of Wadsworth)

pH	turbidity	predicted E. coli	95% CI
6.00	1.00	25.88	(6.98, 95.91)
6.00	3.39	78.37	(23.54 , 260.93)
6.00	**11.47**	**237.36**	**(70.70 , 796.85)**
6.00	**38.86**	**718.87**	**(189.59 , 2420)**
7.00	1.00	9.43	(4.19 , 21.21)
7.00	3.39	28.57	(15.37 , 53.09)
7.00	11.47	86.52	(45.93 , 162.97)
7.00	**38.86**	**262.02**	**(112.99 , 607.63)**
8.00	1.00	3.44	(1.87 , 6.31)
8.00	3.39	10.41	(7.69 , 14.11)
8.00	11.47	31.53	(22.77 , 43.67)
8.00	38.86	95.51	(50.37 , 181.08)
9.00	1.00	1.25	(0.50 , 3.15)
9.00	3.39	3.80	(1.78 , 8.08)
9.00	11.47	11.49	(5.36 , 24.64)
9.00	38.86	34.81	(13.61 , 89.03)

Predicted E. coli counts (east of Wadsworth)

Samples taken at bank

pH	turbidity	predicted E. coli	95% CI
6.00	**1.00**	**841.74**	**(216.77 , 2420)**
6.00	**3.39**	**309.09**	**(162.02 , 589.65)**
6.00	11.47	113.50	(46.32 , 278.09)
6.00	38.86	41.68	(7.38 , 235.43)
7.00	**1.00**	**158.29**	**(73.94 , 338.84)**
7.00	3.39	122.30	(82.40 , 181.53)
7.00	11.47	94.50	(60.59 , 147.39)
7.00	38.86	73.02	(31.55, 168.98)
8.00	1.00	29.77	(17.66 , 50.17)
8.00	3.39	48.40	(32.96 , 71.07)
8.00	11.47	78.69	(56.51 , 109.56)
8.00	**38.86**	**127.94**	**(85.94 , 190.47)**
9.00	1.00	5.60	(2.14 , 14.61)
9.00	3.39	19.15	(10.24 , 35.82)
9.00	11.47	65.52	(31.44 , 136.55)
9.00	**38.86**	**224.17**	**(69.81 , 719.91)**

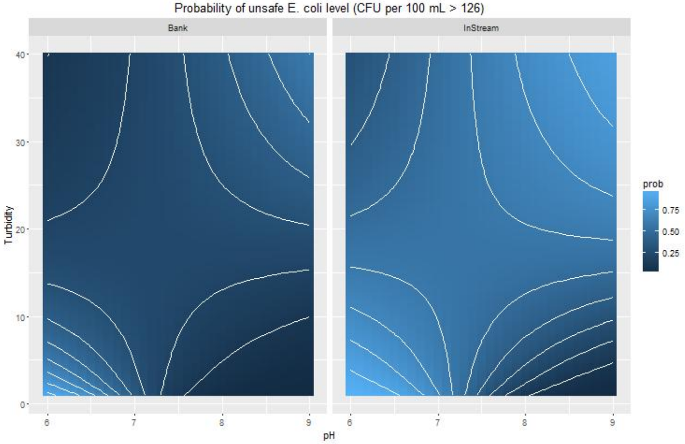
Predicted E. coli counts (east of Wadsworth)

Samples taken in stream

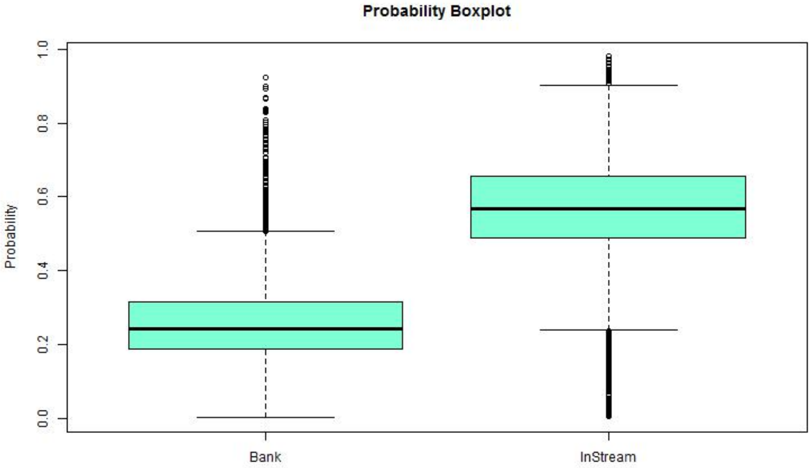
pH	turbidity	predicted E. coli	95% CI
6.00	**1.00**	**1691.12**	**(476.97 , 2420)**
6.00	**3.39**	**620.97**	**(323.56 , 1191.78)**
6.00	**11.47**	**228.02**	**(81.55 , 637.54)**
6.00	38.86	83.73	(12.90 ,543.48)
7.00	**1.00**	**318.01**	**(172.53 , 586.16)**
7.00	**3.39**	**245.72**	**(179.40 , 336.56)**
7.00	**11.47**	**189.86**	**(110.92 , 324.99)**
7.00	**38.86**	**146.70**	**(55.53 , 387.57)**
8.00	1.00	59.80	(43.56 , 82.09)
8.00	3.39	97.23	(82.78 , 114.20)
8.00	**11.47**	**158.09**	**(130.00 , 192.25)**
8.00	**38.86**	**257.05**	**(177.39 , 372.47)**
9.00	1.00	11.25	(4.65 , 27.21)
9.00	3.39	38.47	(24.45 , 60.55)
9.00	**11.47**	**131.64**	**(75.76 , 228.73)**
9.00	**38.86**	**450.38**	**(159.32 , 1273.17)**

Logistic Model

We changed the response from being a numeric variable to a binary variable dependent on safety. This now will output the probability of unsafe exposure to E. coli.

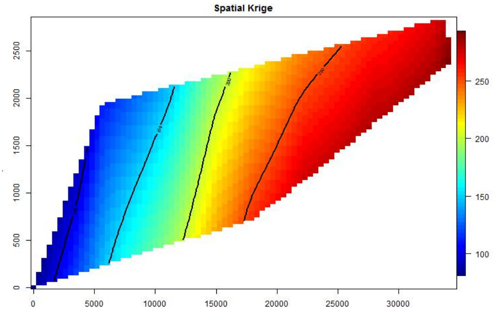


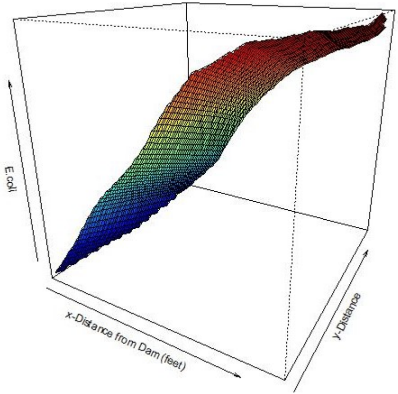
Logistic Model



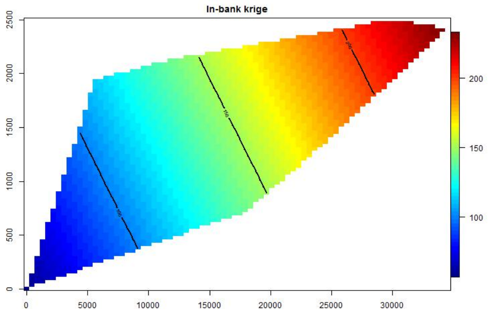
Spatial Kriging

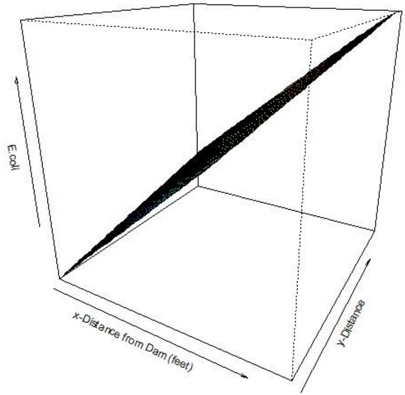
Another method in approaching this type of problem is spatial kriging which spatially interpolates between testing locations. A time dimension can be added to see effects over time.





Bank Results

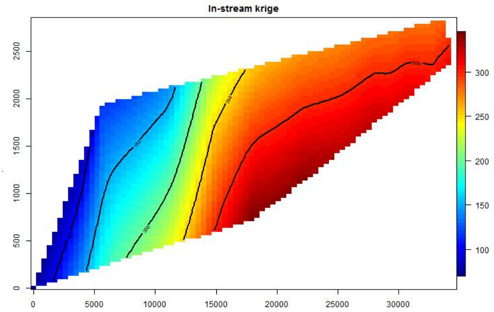


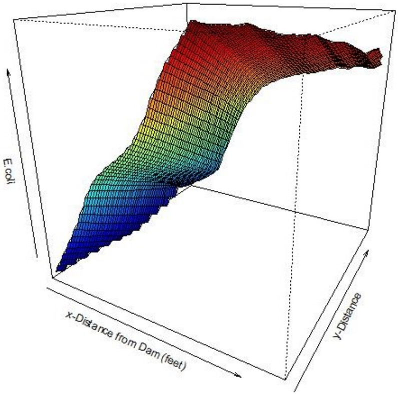


Note:

- Steady increase as we move East
- Surpass safety levels around Wadsworth/Sheridan
- Reach maximum of around 3-times the safety threshold in Englewood

InStream Results





Note:

- Higher E.coli counts
- Surpass safety levels closer to the Dam
- Clear spike in levels around Wadsworth/Sheridan

Metropolitan State University of Denver

Temporal consideration

Synopsis of Results

Each model has come to a conclusion:

Multiple Regression

Our variables account for a very low percentage of E. coli variability; perhaps more is contributed by humans?

Logistic Regression

The probability of being over the threshold of safety is significantly higher in-stream regardless of pH and turbidity. For the bank, the probability increases the most when both pH and turbidity are low or high.

Spatial Kriging

The further east you are in Bear Creek, the more E. coli you will find. It is slightly safer on the bank of the river than in-stream but east of Wadsworth is dangerous regardless.

Future work

As we worked, we discovered further topics that would be interesting to investigate:

- A full spatio-temporal krige
- One-way direction kriging (Thanks Matt Pocernich!)
- Look further in weather effects
- Look into spatial/temporal dependency in the regression models
- Continue to sample and investigate other variables

Credits

We would like to thank:

- Rachel Hansgen and Groundwork Denver for raising awareness of this issue!
- Dr. Ferrell and the students of the Biology department for gather data!
- Dr. Ribble for so much support and guidance with statistics and programming in R!

Sources

* Groundwork Denver [Bear Creek Project]
* H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
* D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf
* Fields Development Team (2006). fields: Tools for Spatial Data. National Center for Atmospheric Research, Boulder, CO. URL http://www.cgd.ucar.edu/Software/Fields.