Towards Mapping Industrial Activity to Chemical Contaminants

Fostiropoulos Iordanis[1,*], Lan Ouyu[1,*], Lapsiya Purvak[1,*], Zhao Eric[2,†]

1 University of Southern California
2 California Institute of Technology

***{fostirop,olan,lapsiya}@usc.edu**
†elzhao@caltech.edu

Abstract

Industrial activity can directly influence the water quality for a given geographical location. Depending on available natural resources in a geographical location, topography, legislation and zoning regulation, different counties within the US can be attractive to different industries. Different industries produce as byproducts different chemicals for their operational processes. We correlate the chemical contamination footprint of each industry to water supply contaminants. We use data from different counties and correlate industrial activity with introduction to specific chemical specimens in water supply.

Introduction

Industrial activity changes from county to county. It can be a difficult task to track the extend at which an industry operates in a given county.

Economical data such as employment numbers from that industry can sometimes not be sufficient to determine the level of activity. Companies within that industry can have multiple locations and divide their operational processes to different places. Such companies while operating within an industry, e.g. Mining Industry, might not have an active mining operation in a given location due to lack of mining resources but still operate part of their processes there, such as processing, R&D and more due to advantageous electricity costs, available human resources and more.

Tracking the impact an industry has on a region then becomes a challenge. Since we are unable to identify the activities that take place from the data publicly available. Moreover different processes within an industry have different byproducts. Many of which can end up in the water supply either unintentionally or as waste of their processes. Contamination then can be a result of bad waste management or bad safety protocols. To add to that, industries can transfer their processes based on changes in legislation, zoning requirements, and more. For example a company might determine it is more profitable to transfer production of Chemical A to plant B from Plant A. Such changes in manufacturing processes would be not possible to track only based on economic factors and employment data.

Changes then in water contamination data can seem stochastic because we lack the context that impacts them. Determining the extend at which an industry operates in a given location is then critical to understanding the impact of that industry to the water quality.

Water usage data alone for specific industries on the other hand can be misleading in determining by itself industrial activity for that county. We determined a negative correlation between industrial activity

Methodology

Water usage is an important factor in industrial activity. However, we are only given one year (2010) statistical information on water usage. To handle this, the first step is to build a regression model from historical education attainment and the population in 2010 to water usage. Then the water usage of other years could be inferred from this regression model. Furthermore, industrial occupation and water usage vectors are concatenated to represent the industrial features. Specifically, it is normalized every feature to a mean and variance of 1.

We build a profile of each county for a given year by concatenating a vector of the county's water usage in 2010 and their occupation statistics for the given year. We normalized the water usage data for population and droughts during 2010. We then used regression to determine water usage based on population and drought, and subtracted the regression values from the water usage data. We are then able to determine the baseline measurements for that county's industrial activity in 2010.

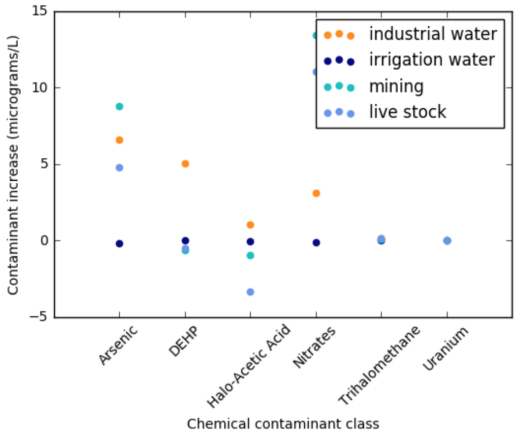
We use the industrial activity data and the chemical contamination levels to determine the extend of contamination in the US for each specific industry.

Analysis

Given the effective representation of industry activity of each county, we try to figure out the relationship between industrial activity and contaminant levels. Firstly, a regression model is fitted from industrial features to each contaminant levels. Then the contaminant level could be predicted by the regression model after perturbing. In order to figure out the impact significance of different contaminants, chemicals are clustered based on their regression coefficients.

In addition to the influence of industrial activities on each chemical contaminants, we also care about the impact of the year-wise change of industrial activities.

Results



For each industry the results.

References