Capstone Proposal 1

Time series analysis and forecast of Benzene levels in gasoline

Introduction

Gasoline releases from the underground storage tanks are frequent causes of groundwater contamination. The nature and magnitude of contamination from such releases depend on composition of the gasoline. The gasoline composition in US varies spatially as well as temporally for various reasons which include crude oil sources, refinery equipment and capabilities, fuel performance characteristics, industry standards and regulation.

Federal regulations have mandated limits on certain component concentrations and properties. For example, limits of benzene, ethers and alcohols concentrations, and vapor pressure are set by these regulations. These components are primarily regulated for their impact on air pollution. However, components like benzene and other aromatics, methyl *tert*-butyl ether (MTBE) and other oxygenates are known ground water contaminants and more importantly, injurious to human health.

The Clean Air Act Amendments (CAAA) of 1990 (42 U.S. Code 4701) expanded regulation of fuels initially set by the Clean Air Act 1970.The CAAA introduced several requirements that have had a major impact on gasoline composition throughout the United States, beginning with implementation in 1992 and 1995, and continuing to the present. The most important requirements were the total ban on lead in gasoline, and new requirements for three types of gasoline: conventional (CG), reformulated (RFG), and oxygenated (OG). Both reformulated gasoline (RFG) and oxygenated gasoline (OG) required oxygen-containing additives to facilitate clean burning of fuel.

The gasoline release events are usually not discovered immediately. When discovered, it is often very expensive to measure the level of contaminants. They are derived by physics based modeling (advection-dispersion models) of the groundwater and the contaminants. Such modeling is highly dependent on the original concentration of components in gasoline. Since, the concentrations of these contaminants vary spatially and temporally, it would beneficial to analyze and build a time series model of the contaminant concentration. This would allow to get a precise estimate of the concentration in the past as well as to forecast it into the future. Such models will not only aid in the physical process based modeling, it will also help determine the effectiveness of the regulations in achieving their objectives. Thus main objectives of this project are:

1. To analyze the contaminant concentration in gasoline to examine the effectiveness of the regulations.
2. To build a time series model to estimate and forecast contaminant concentration

Methodology

This study analyzes concentration of a specific contaminant, benzene in a specific regulation zone that consists of New York, New Jersey and Connecticut. The RFG program was implemented in this zone and it limited the amount of benzene and total aromatics in reformulated gasoline. The limits were met either on per-gallon basis(<1.00 % vol) or averaged basis (<1.3 % vol). These limits were enforced starting 1995. The benzene content in all US gasoline was reduced to 0.62 % vol in 2011 by the Mobile Sources Air Toxics Rule.

A long-term data between 1976 to 2017 has been compiled from three different sources and is available as csv files. The dataset contains concentration of measurement of different gasoline components acquired at different points of sale throughout the regulation zone. However, the analysis will be limited to benzene concentrations measured from 1995 to 2017.

Three different time series forecasting methods will be used to construct the model, Simple Exponential Smoothing (SES), Holts linear trend method and Seasonal Auto Regressive Integrated Moving Average (SARIMAX) model. The data between 1995 and 2010 will be used to train the models and data after 2011 will be held back to test the models. All the modeling will be done using python and it's relevant packages. The jupyter notebook will be used for the coding and presenting results.

Potential beneficiaries

The study will be beneficial to federal agencies to review the effectiveness of regulations. It will also be useful to researchers studying the Leaking underground storage tanks and groundwater contaminations. The study will also be of interest to general public.

Deliverables

Codes used for data exploration and model development
Report presenting the analysis, results and conclusions on models.